NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

ANALYSIS OF INTERNET TELEPHONY AND THE H.323 MULTIMEDIA STANDARD

by

Georgios Floros

September 1999

Thesis Advisor: Gilbert M. Lundy
Second Reader: Geoffrey G. Xie

Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE September 1999	3. REPORT TYPE AND DATES COVERED Master's Thesis

4. TITLE AND SUBTITLE ANALYSIS OF INTERNET TELEPHONY AND THE H.323 MULTIMEDIA STANDARD	5. FUNDING NUMBERS
6. AUTHOR(S) Lt Georgios Floros, HN	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT	12b. DISTRIBUTION CODE
Approved for public release; distribution is unlimited.	

13. ABSTRACT *(maximum 200 words)*

Internet Telephony is an emerging and promising technology which enables the transport of voice over data networks. If it is going to be succesful, standardization will be critical. The purpose of this study is to assess the suitability of the H.323 standard for establishing and performing real-time voice conversations over IP networks. The goals of the study were to (1) examine the current status of Internet Telephony, (2) conduct research on the current Internet Telephony software solutions in terms of quality, performance, interoperability and H.323 compliance, (3) analyze and evaluate the H.323 standard, (4) compare H.323 with Session Initiation Protocol, and (5) provide recommendations for further improvements of H.323. The study shows the complexity of H.323 and highlights the areas where more considerations are required. Part of the study includes the testing of 10 Internet Telephony programs. The tests show that H.323 compliance does not guarantee interoperability and voice quality. Also it is shown that the standard is not yet mature despite its popularity. However, it is assessed that Internet Telephony is a technology which will experience tremendous expansion during the forthcoming years. Based on the analysis and evaluation results, recommendations are provided in order for the H.323 to be more suitable for Internet Telephony.

14. SUBJECT TERMS	15. NUMBER OF PAGES
H.323 standard, Internet Telephony, Voice over IP, Multimedia Communication	128
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFI- CATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

Approved for public release; distribution is unlimited

ANALYSIS OF INTERNET TELEPHONY AND THE H.323 MULTIMEDIA STANDARD

Georgios Floros
Lieutenant, Hellenic Navy
B.S., Hellenic Naval Academy, 1988

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

from the

NAVAL POSTGRADUATE SCHOOL
September 1999

Author: ________________________
Georgios Floros

Approved by: ________________________
Gilbert M. Lundy, Thesis Advisor

____
Geoffrey G. Xie, Second Reader

____
Dan C. Boger, Chair
Department of Computer Science

ABSTRACT

Internet Telephony is an emerging and promising technology which enables the transport of voice over data networks. If it is going to be succesful, standardization will be critical. The purpose of this study is to assess the suitability of the H.323 standard for establishing and performing real-time voice conversations over IP networks. The goals of the study were to (1) examine the current status of Internet Telephony, (2) conduct research on the current Internet Telephony software solutions in terms of quality, performance, interoperability and H.323 compliance, (3) analyze and evaluate the H.323 standard, (4) compare H.323 with Session Initiation Protocol, and (5) provide recommendations for further improvements of H.323. The study shows the complexity of H.323 and highlights the areas where more considerations are required. Part of the study includes the testing of 10 Internet Telephony programs. The tests show that H.323 compliance does not guarantee interoperability and voice quality. Also it is shown that the standard is not yet mature despite its popularity. However, it is assessed that Internet Telephony is a technology which will experience tremendous expansion during the forthcoming years. Based on the analysis and evaluation results, recommendations are provided in order for the H.323 to be more suitable for Internet Telephony.

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ADPCM......... Adaptive Differential PCM

ARPA.......... Advanced Research Project Agency

ASN.1......... Abstract Syntax Notation 1

ATM........... Asynchronous Transfer Mode

B-ISDN........ Broadband ISDN

CELP.......... Code Excited Linear Predictor

CS-ACELP...... Conjugate-Structure Algebraic CELP

CVSD.......... Continuously Variable Scope Delta

DLS........... Dynamic Lookup Service

GK............ GateKeeper

GSM........... Global Systems for Mobile Communications

HTTP.......... HyperText Transfer Protocol

IETF.......... Internet Engineering Task Force

IP............ Internet Protocol

ISDN.......... Integrated Services Digital Network

ITU........... International Telecommunication Union

Kbps.......... Kilobits per second

LAN........... Local Area Network

LPC........... Linear Predictive Vocoder

MC............ Multipoint Controller

MCU........... Multipoint Control Unit

MGCP.......... Media Gateway Control Protocol

MMUSIC........ Multiparty Multimedia Session Control

Modem......... Modulator/demodulator

MP............ Multipoint Processor

PBN........... Packet Based Network

PBX........... Private Branch eXchange

PCM........... Pulse Code Modulation

PER........... Packed Encoding Rules

POTS.......... Plain Old Telephone Service

PSTN.......... Public Switched Telephone Network

QoS........... Quality of Service

RAS........... Registration Admission and Status

RFC........... Request For Comment

RPE........... Regular Pulse Excited

RSVP.......... Resource ReserVation Protocol

RTCP.......... Real-time Transport Control Protocol

RTP........... Real-time Transport Protocol

RTSP.......... Real Time Streaming Protocol

RTT........... Round Trip Time

SCN........... Switched Circuit Network

SIP........... Session Initiation Protocol

TCP........... Transmission Control Protocol

UDP........... User Datagram Protocol

VoIP.......... Voice over IP

WAN........... Wide Area Network

ACKNOWLEDGEMENTS

I would like to express acknowledgment and appreciation to my thesis advisor Dr. Gilbert Lundy for his assistance, direction, and dedication throughout my course of study. Special thanks also go to Professor Geoffrey Xie for his insights on the reviewing of my thesis.

I express my gratitude to my parents for providing me with the best education and moral standards they could in the early stages of my life and to the Hellenic Navy for giving me the opportunity to enhance my knowledge by attending NPS.

Acknowledgment and appreciation is due to my friend and colleague Michail Demestihas for his assistance and inputs during the software testing. Appreciation is also due to Bonnie.

Lastly I want to thank my beloved wife Elena not only for her love, patience, encouragement and support throughout the duration of my studies, but also for her actual invaluable contribution during the software testing, and the editing of this thesis.

I. INTRODUCTION

A. OVERVIEW

Throughout this century all the advanced Navies of the world have shared one common quality: In addition to their good war-fighting capabilities, they have pioneered the development and use of technology to achieve the impossible. As the century draws to a close, the pioneering effort must continue. New challenges lie ahead. The information revolution has created new technologies and opportunities, which alter the way information is acquired. One emerging and promising technology is the integration of voice and data over existing computer networks and the Internet. This technology is also known as **Internet Telephony**.

Data networks have progressed to the point that it is now possible to support multimedia applications over the existing enterprise and military networks and the Internet. New QoS (Quality of Service) supporting protocols, such as Ipv6, differentiated services in Ipv4, and Real-time Transport Protocol (RTP), are now entering the enterprise network. Given this extensive deployment of data networking resources, the following question naturally presents

itself: Is it possible to use the investment already made to carry real-time voice in addition to the data?

Internet phones and Internet videophones are entering the market, making it possible to talk to and see one or more remote parties over the Internet. However until very recently, Internet telephone solutions have tended to be incompatible (i.e., one could communicate with an Internet phone user only if he or she happened to be using the same software). While there are still several proprietary solutions, the emergence of the H.323 and recently the SIP standards has the potential to provide a foundation for audio, video, and data communications across the Internet.

Besides the potential for savings on long-distance phone charges, Internet phones already have a place in the business and military world. The technology is good for linking military forces with the transmission of voice, video and data from a single desktop PC. Using laptop computers and satellite links, a distant commander can communicate with the scene of action commander using only one channel for voice, video and data.

However, the use of the existing IP routers for voice communications is not an easy task. Routers work well for traditional data applications, but new broadband multimedia

applications need different forwarding treatment, higher throughput, and tighter QoS control. The traditional network service on the Internet is best-effort packet transmission. In this service there is no guarantee of delivery. To support real-time services in an IP environment, the Resource Reservation Protocol (RSVP) has recently been advanced as the signaling protocol to enable network resources to be reserved for a connectionless data stream. To support QoS-sensitive applications, such as voice and video, Intranets and the Internet need to provide differentiated quality of service. To make QoS on IP networks a reality, new specifications and standards need to be implemented.

While a number of vendors tout products that support voice over the Internet, performance issues remain. Particularly, the issues of latency and audio signal quality need to be addressed. If voice over IP is going to be successful, standardization will be critical. The H.323 standard, developed by the International Telecommunication Union (ITU-T), describes terminals, equipment, and services for multimedia communication over LAN and IP networks that do not provide a guaranteed quality of service. Support for voice is mandatory in the standard, while data and video

are optional; but if supported, the ability to use a specified common mode of operation is required, so that all terminals supporting that media type can interwork.

B. BACKGROUND

The concept of integrating voice and data is dated back in 1975 when ARPA with project DACH-15-75-C0135 funded some researchers to look at the feasibility *of integrated voice and data packet networks*. A few years later the research for ISDN started in Japan. However the mainstream work was focused on carrying voice and data over circuit-switched TDM networks and the term "Internet" had not even introduced yet. The first actual Internet Telephony application was built in 1994 by some savvy Israelis who wrote software that let them make voice phone calls over the Internet. The motivation for them was high because the Israeli Phone Company, *Bezek*, charged $2.50 a minute to call New York. Vocaltec then built the first commercial product in 1995 and it was called "Internet Phone".

The need for standardization of Internet Telephony was apparent and for that reason the ITU commenced studies the same year for the development of the H.323 recommendation. In 1996 the H.323 standard was approved by the ITU.

Currently the standard is in its second version as new elements and functionality are being added continuously. Beside the ITU's H.323, another standard has begun to emerge as a product of the IETF MMUSIC (Multiparty Multimedia Session Control) working group. It is called SIP (Session Initiation Protocol) which is more Internet-centric and was proposed in February of 1999 with RFC 2543.

C. PROBLEM STATEMENT

Given the current unstable state of Internet Telephony and the numerous problems that need to be addressed and solved, a study was conducted for this thesis on the Internet Telephony and the H.323 standard in particular. The goals of this study were:

- Examine the current status of Internet Telephony with focus on the transmission of Voice over IP networks and how this can be standardized.
- Evaluate the current Internet Telephony software solutions in terms of their quality, performance, interoperability and H.323 compliance.
- Analyze and evaluate the H.323 in order to determine if it is the proper solution for carrying audio,

video and data communications across IP-based networks.

- Compare the H.323 with the emerging SIP and highlight the major differences.
- Provide recommendations for further improvements of the H.323 standard.

D. OBJECTIVE

The purpose of this study is to assess the suitability of the H.323 standard for establishing and performing real-time voice conversations over IP networks. The long-term objective is to contribute in the enhancement and improvement of the existing Internet Telephony technology.

E. SCOPE AND LIMITATIONS

To narrow the scope of this thesis only the technological issues of carrying voice over IP will be presented. The concepts of carrying voice over ATM or Frame Relay will not be addressed. The software will be analyzed only in terms of quality (compared with a standard circuit-switched telephone conversation), interoperability and H.323 compliance. Design issues and complexity will not be addressed. The primary focus will be on the H.323 standard since the SIP was introduced and proposed only recently.

II. STATUS OF INTERNET TELEPHONY

In this chapter we will describe the problems associated with Internet Telephony. We will show the current trends in the Internet Telephony industry and we will discuss the Internet Telephony constraints. The chapter ends with a brief presentation of the expected issues that an Internet Telephony standard should be able to address, thus focusing our attention on what we should expect from H.323, which will be discussed in Chapters III and IV.

A. OVERVIEW

A very short definition for Internet Telephony could be: "The use of the Internet for our telephone needs". This can be achieved by compressing voice and carrying it over data networks. It usually involves IP encapsulation (thus the term Voice over IP is very often used) but other variants are also possible (i.e., voice over ATM and Frame Relay). As has been stated in the introduction, this thesis will address the issues with regard to telephony over IP Networks. Research for Voice over IP/Internet is currently developing in three directions:

1. PC-to-PC

In this model individuals talk online through their PCs. The operation of such a system commences with the connection establishment between two calling parties. After connection establishment the caller talks into a microphone. The microphone is in turn connected to a sound

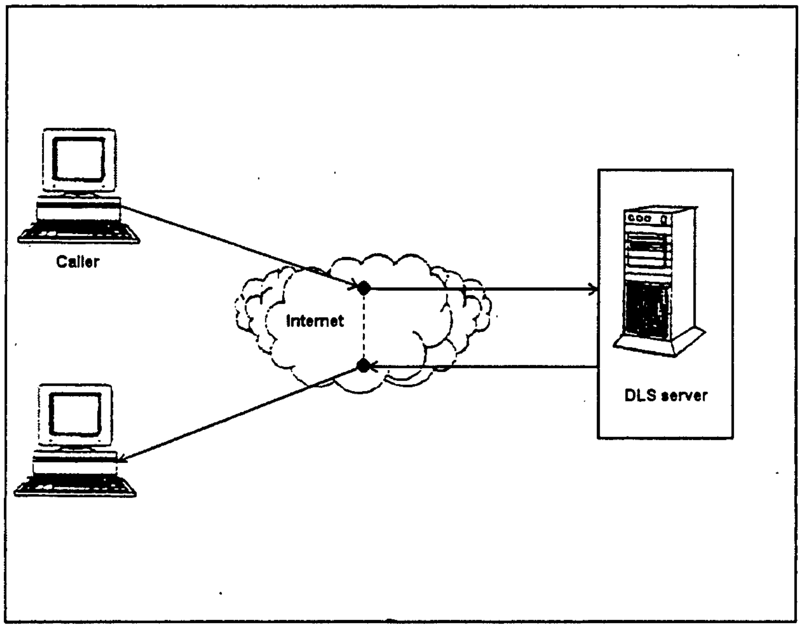


Figure 1: PC-to-PC

card installed in the computer, which accepts an analog waveform and converts it into a digital data stream. Internet telephony software operating on the computer takes the digitized voice data stream, which normally represents a 64-Kbps PCM or a 32-Kbps ADPCM-encoded voice, and compresses the data stream using a proprietary or standardized voice-compression technique. Once this is

accomplished, the software packages the compressed data stream into IP packets for transmission over the Internet. Most PC-to-PC products were originally developed primarily to support modem connections; however, modern products also support LAN-based operations when the LAN is connected to the Internet. This is the simplest form of an Internet Telephony system implementation. The connection can be established with the use of one's e-mail or IP address. For the sake of simplicity, and due to the fact that many Internet users have dynamic IP addresses, several vendors offer of the solution of DLS (Dynamic Lookup service) servers. These servers map a user's current IP address to his/her e-mail so that he or she can always call people using their e-mails as long as they have registered with a DLS server during their current Internet logon session.

2. PC-to-Phone

This scheme allows individuals to make and receive voice calls and messages from/to Public Switch Telephone Network (PSTN) while on the Internet. This approach requires the implementation of a system called "Gateway" which integrates hardware and software and provides the interface between PSTN and the Internet or an Intranet. Gateways generally use both PC-bus compatible interface

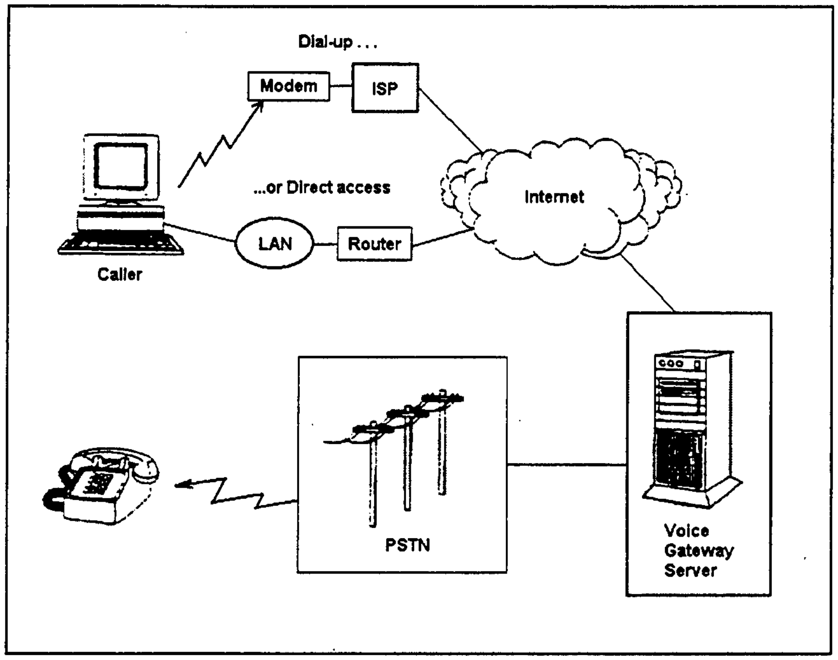


Figure 2: PC-to-Phone

cards, (which terminate the PBX, compress the voice and pack it into IP Datagrams), and a suite of call management software. Each gateway typically receives a telephone number and an IP address, which are entered into the gateway's database. The database provides the mapping of the gateway telephone number to the appropriate IP address.

3. Phone-to-Phone

This is relatively the most complex approach but also the most transparent to the end-user. It allows a person's existing telephone system to be used for calling persons via the Internet. This is achieved by connecting PBXs at each organization location to equipment that behaves in a manner similar to an analog trunk. That is, equipment

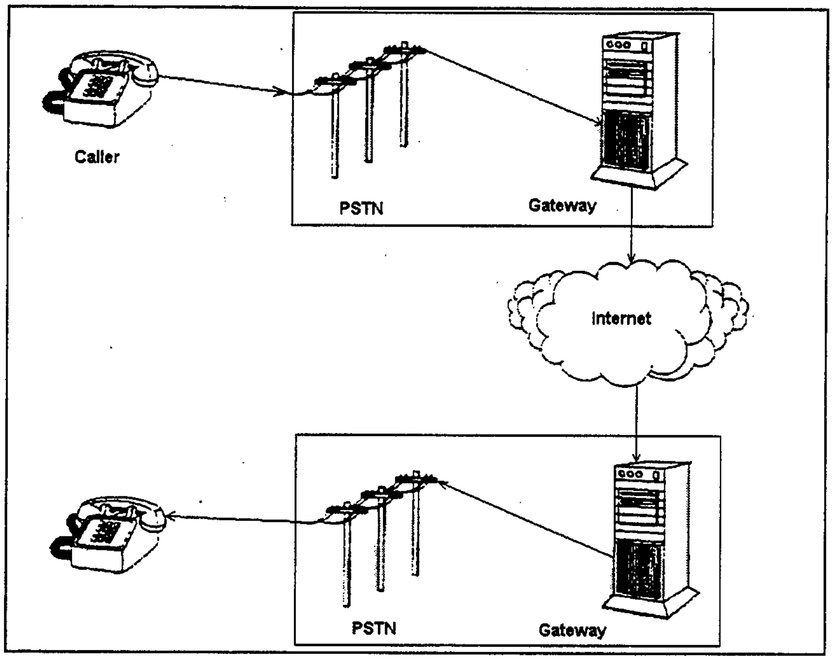


Figure 3: Phone-to-Phone

presents a ring voltage when a call is received, responds to DTMF and/or rotary dialing, and passes caller ID data from an incoming call. In addition, the equipment presents each PBX with call-progress tones such as ringback and busy when outbound calls are made. This equipment is called a "black-box". Access to the black box from the local PBX is accomplished by the caller dialing a predefined prefix and then the extension of the called party. The analog call is routed to the black box, where it is digitized and compressed. In addition, the black box provides appropriate addressing for each packet so that packets are routed to a similar black box at the called location. At that location,

voice-digitized packets are converted back into their original analog form for routing via the destination PBX to the called party.

Some authors prefer to differentiate the above techniques from VoIP. Others like Gilbert Held (1998) use a different term for the Phone-to-Phone approach namely the term "Telephony over the Internet". In this thesis however we will use the term Internet Telephony to describe all the different aspects of this industry and the emphasis will be placed on the first two techniques mentioned above.

B. INTERNET TELEPHONY CONSTRAINTS

There are numerous constraints and problems associated with the implementation of an Internet Telephony system. The lack of a uniform standard approach forced vendors to develop their own proprietary solutions for their products. Therefore interoperability between products has become difficult if not impossible.

The main factor that affects all the Internet Telephony applications is the Quality of Service (QoS). This is just a collection of parameters that relate to a sequence as seen at the source and as seen at the destination. These parameters include *Bandwidth, Delay,*

Delay Jitter, and Packet or Cell Loss Probability, and they are summarized in table 1. Since we assume that the underlying network for Telephony will be an IP based one, then clearly we cannot have a hardware guarantee. A software guarantee is needed instead and that's where standards come into place. With "guarantee" we mean the required resources that need to be reserved by the system well ahead of the start of a conference. In practice this is hard especially with Delay and Delay Jitter. These consist of propagation delay, transmission time, queuing delays, and protocol overheads. In other words they consist of factors that are difficult to be quantified. Needless to say that if a system fails to address these problems adequately then the quality of transmitted voice will be affected heavily with occurrences of voice clipping, latency and distortion.

Table 1: QoS Parameters

Name	Description	Allowed Values
Bandwidth	The capacity of the transfer mechanism that is available between source and destination.	14.4 Kbps (Dial-UP) Up to 100 Mbps (Fast Ethernet)
Delay	The time a unit of information	10 ms (Normal

	(packet, cell, or bits) spends in the transmission system.	Phone)
Delay Jitter	The variation of delay.	40 - 60 ms (buffer)
Packet Loss	The ratio of units of information that the application can afford to lose.	Depends on CODEC (may be up to 50%)

1. Bandwidth and Speech Compression Algorithms

This is perhaps the only constraint that has been addressed effectively with regards to proper standardization. Many algorithms have been implemented which perform extended compression thus eliminating the amount of bandwidth needed. However the ability to reproduce a natural-sounding conversation requires a trade-off between the speech-coding scheme and processing power of the PC. For the composition of speech three general types of encoding exist. The first is known as ***waveform encoding*** and it is based on the Sampling Theorem and the second is called *voice coding* or ***vocoding.*** As a combination of these two methods a third method was developed based on waveform and vocoding-a hybrid technique which was

naturally called ***hybrid coding.*** The most important coding techniques are explained below.

a) Pulse Code Modulation (PCM)

This is by far the most commonly used method of waveform encoding on a worldwide basis and it was developed during the 1960s. It is based on a three-part process: sampling, quantization, and coding. Under PCM, an analog signal is sampled 8000 times per second. This sampling rate is based on the Nyquist theorem, which dictates that the number of sample points must be at least twice the maximum frequency of the signal for the latter to be reconstructed with the maximum fidelity. The standard voice channel is filtered to produce a bandwidth of 3000 Hz (3 - 3.3 kHz). However the filters do not work instantaneously and allow some lower power speech to pass below 300 Hz and beyond 3300 Hz thus making the bandwidth approximately equal to 4000 Hz which enforces the selection of a sampling rate of 8000 samples/sec. The sample signal is then quantized resulting in the coding of each sample with 8-bit bytes. The entire process results in a digital data stream of **64 Kbps**. PCM was standardized by the ITU as Recommendation **G.711** and offers near-toll quality voice reproduction.

b) Adaptive Differential PCM

Due to the fact that the human speech tends to repeat waveforms because of the vibration of the vocal cords a new coding scheme emerged. It was based on the prediction of samples if the error between the predicted samples and actual speech samples have a lower variance than the original speech samples. If so, the difference between the actual sample and the predicted sample could be quantized using fewer bits than the original speech sample. This technique forms the basis for a series of differential PCM methods, including *adaptive differential PCM* (ADPCM), in which the predictor and quantizer adaptively adjust to the characteristics of speech being coded. ADPCM was standardized by the ITU in the mid-1980s as Recommendation **G.721** and resulted in operating rate of **32 Kbps.**

c) Continuously Variable Slope Delta Modulation

In the continuously variable slope delta (CVSD) modulation technique, the analog input voltage is compared to a reference voltage. If the input is greater than the reference, a binary 1 is encoded, while a binary 0 is encoded if the input voltage is less than the reference level. Most CVSD systems sample the input at 32,000 or

16,000 times per second, resulting in a bit rate of **32** or **16 Kbps** representing a digitized voice signal. Another popular CVSD rate is **24 Kbps**, which results from a sampling rate of 24,000 times per second.

d) CELP and CS-ACELP

The Code Excited Linear Predictor (CELP) speech coder is a hybrid scheme that employs both waveform and vocoding techniques resulting in an analysis-by-synthesis process to code speech. Several versions have been implemented with the most important being the FS -1016 (DOD standard) and the G.728 recommendation. A variation of CELP is the G.729 recommendation also known as Conjugate-Structure Algebraic CELP but the most interesting is the G.723.1 one. A G.723.1 coder supports both 5.3 and 6.3 Kbps rates with the higher bit rate providing a higher quality of reproduced voice. A silence suppression capability is included, which is the product of a voice activity detection technique and can result in an average rate of 2.65 - 3.15 Kbps. The G.723.1 besides the low data rate is also relatively more robust when it comes to the handling of lost packets (a weakness of IP networks). This enhancement is also referred to as frame erasure ability

and allows some packets to be discarded if they are late, thus avoiding latency in conversations.

Table 2: Speech Compression Algorithms

Recommendation or Standard	CODEC	Type	Data Rate (Kbps)	Delay (ms)
G.711	PCM	Waveform	64	0.125
G.721	ADPCM	Waveform	32	0.125
-	CVSD	Waveform	24	0.625
FS - 1015	LPC	Vocoder	2.4 - 4.8	22.5
GSM	RPE	Hybrid	13	20
FS - 1016	CELP	Hybrid	4.8	15
G.728	CELP	Hybrid	16	0.625
G.729	CS-ACELP	Hybrid	4 - 8	15
G.723.1	CS-ACELP	Hybrid	5.3 and 6.3	37.5

At a first glance it seems that the most suitable compression scheme is the one offered by G.723.1. It has been shown *(Held, 1998)* that the G.723.1 coders were rated at 3.98 out of a scale of 4.00 thus pending only 0.02 points from full toll quality. Indeed the G.723.1 has gained significant popularity and it is also included in the H.323 standard, as it will be shown later. However the choice of the right coding scheme involves several other

considerations as well. G.723.1 for example adds significant delay overhead due to the complexity of the coding algorithm. On the other hand we must consider the size of the packet that each compression algorithm requires. Large packet sizes increase the delay. G.711 for example requires a long packet length and drains a large amount of bandwidth.

2. Real Time Data Transport Mechanisms

The flow of Real-time voice and video streams have some specific requirements that differ from "traditional" Internet data services. These requirements are mainly the following:

Sequencing: The packets must be re-ordered in real time at the receiver, when they arrive out of order. If a packet is lost, it must be detected and compensated for without retransmissions.

Synchronization: When different types of media are being used in a session, there must be a means to synchronize them. (i.e., audio must be matched with video).

Payload identification: In the world of the Internet it is necessary for the applications to be able to change the encoding for the media (payload) in order to adjust to changing bandwidth availability or for "ad-hoc" situations.

A proper mechanism is needed to identify the encoding for each packet.

Frame indication: Video and audio are sent in frames. It is necessary to indicate to a receiver where the beginning or end of frames is, for the synchronized delivery to higher layers.

The above services are provided in the Internet by the Real Time Transport Protocol (RTP). RTP has two components. The first is RTP itself, and the second is RTCP, the Real Time Control Protocol. RTP is generally used in conjuction with the User Datagram Protocol (UDP), but can make use of any packet based lower-layer protocol.

0		4		8	9	16	32
V	P	X	CC	M	Payload type	Sequence number	
Timestamp							
Synchronization source identifier (SSRC)							
Contributing source identifier (CSRC)							
. . .							
Contributing source identifier (CSRC)							

Figure 4: RTP Header

When a host wishes to send a media packet, it takes the media, formats it for packetization, adds any media-specific packet headers, prepends the RTP header, and places it in a lower-layer payload.

The RTP header (Figure 4) is 12 bytes long. The V field indicates the protocol version. The X flag signals the presence of a header extension between the fixed header and the payload. If the P bit is set, the payload is padded to ensure proper alignment for encryption. Table 3. Summarizes the payload types as defined in RFC 1890.

A random 32-bit synchronization source SSRC identifier distinguishes users within a multicast group. Having an application-layer identifier allows the application to easily distinguish streams coming from the same translator and associate receiver reports with sources. RTP supports the notion of media dependent framing to assist in the reconstruction and playout process. The marker M bit provides information for this purpose. The payload type identifies the media encoding used in the packet. The sequence number increments sequentially from one packet to the next, and is used to detect losses and restore packet order. The timestamp, incremented with the media sampling frequency, indicates when the media frame was generated.

Table 3: Payload Types (RFC 1890)

0	PCMU audio	16-23	unassigned audio
1	1016 audio	24	unassigned video
2	G721 audio	25	CelB video
3	GSM audio	26	JPEG video
4	unassigned audio	27	unassigned

5	DV14 audio (8 kHz)	28	nv video
6	DV14 audio (16 kHz)	29-30	unassigned video
7	LPC audio	31	H261 audio
8	PCMA audio	32	MPV video
9	G722 audio	33	MP2T video
10	L16 audio (stereo)	34-71	unassigned
11	L16 audio (mono)	72-76	reserved
12-13	unassigned audio	77-95	unassigned
14	MPA audio	96-127	dynamic
15	G728 audio		

3. Packet Delay, Jitter, and Loss Handling

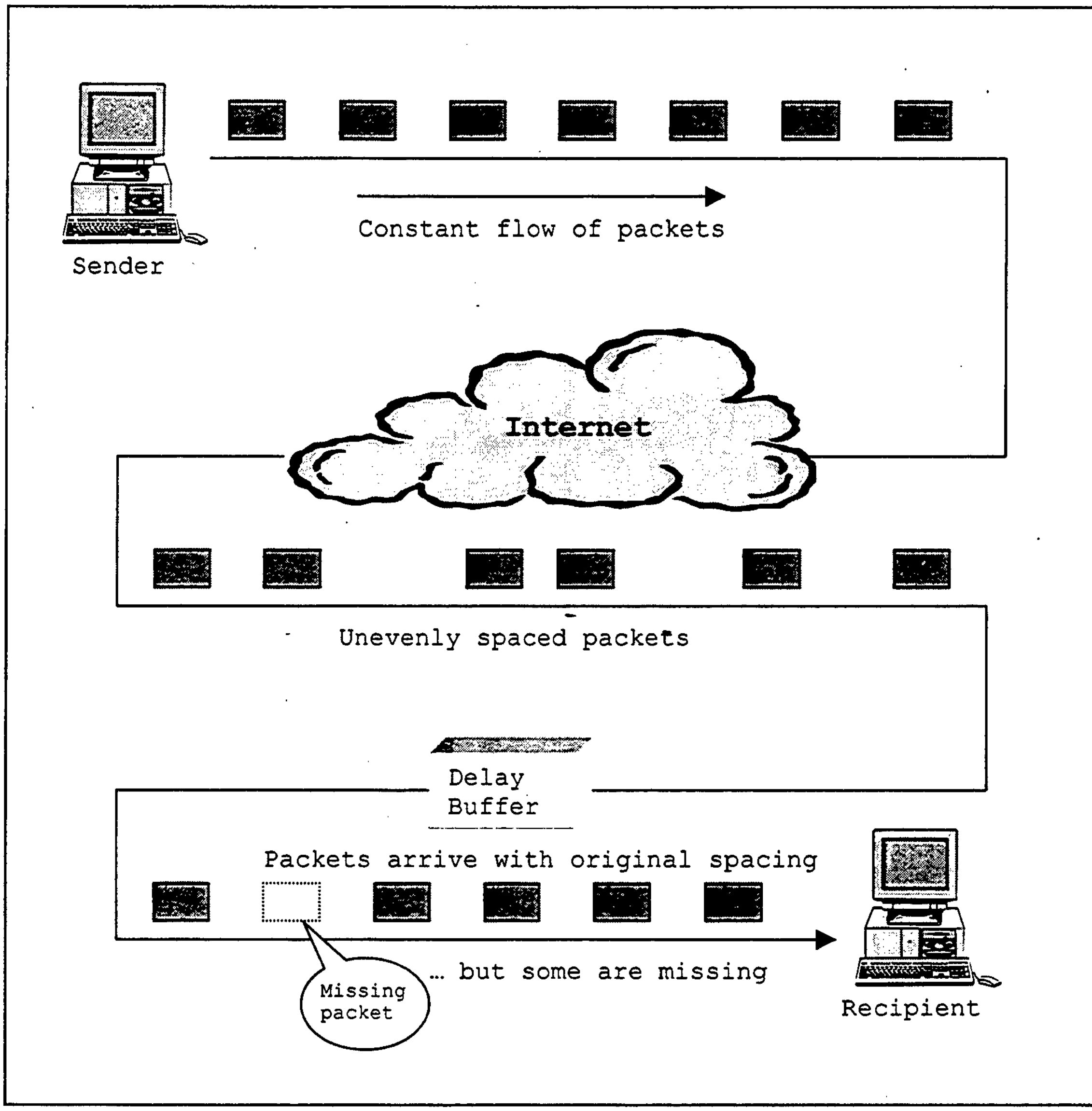


Figure 5: Real Time Traffic Problems

When data is being transferred, a slight delay in the arrival of one or more packets is usually not noticeable. Similarly, the loss of packets as they flow through an IP network resulting from congestion and routers, workstations, or gateway discarding packets is compensated for by the retransmission of discarded packets. However, when packets transport digitized voice, normal data transmission methods cannot be used. This is because the loss or delay of packets results in the disruption of speech intelligibility.

The measurement of the amount of delay or delay jitter can be achieved using the attributes of RTP. Delay can be measured with the use of the RTP Timestamp field in the RTP header.

This field allows recipients to understand the relative time passed between the transmission of two packets and place them in the appropriate time sequence afterwards. Measuring delay jitter is somewhat more complex. Actually there is no simple way to measure this quantity at the receiver, but it is possible to estimate the average jitter in the following way as shown by Stallings (1998). At a particular receiver define the following parameters for a given source:

S(I) = Timestamp from RTP data packet *I*.

R(I) = Time of arrival for RTP data packet *I*, expressed in RTP Timestamp units. The receiver must use the same clock frequency (increment interval) as the source but need not synchronize time values with the source.

D(I) = The difference between the interarrival time at the receiver and the spacing between adjacent RTP data packets leaving the source.

J(I) = Estimated average interarrival jitter up to the receipt of RTP data packet *I*.

The value of D(I) is calculated as:

$$D(I) = (R(I) - R(I-1)) - S(I) - S(I-1)) \qquad (II-1)$$

Thus *D(I)* measures how much the spacing between arriving packets differs from the spacing between transmitted packets. In the absence of jitter, the spacings will be the same and *D(I)* will have a value of 0. The interarrival jitter is calculated continuously as each data packet *I* is received, according to the formula:

$$J(I) = \frac{15}{16} J(I-1) + \frac{1}{16} \left| D(I) \right| \qquad (II-2)$$

J(I) is calculated as an exponential average of observed values of *D(I)*. Only a small weight is given to the most

recent observation, so that temporary fluctuations do not invalidate the estimate. The jitter measure may provide a warning of increasing congestion before it leads to packet loss.

There exist a few methods that can be used to compensate for the loss or delay of packets at the application level. Those methods involve repairing lost or delayed packets with periods of silence or with synthetic speech.

Buffering: Incoming packets are buffered and slightly delayed. Then they are released at a constant rate to the software that generates the audio. However if buffering time is less than delay time of a packet, then this packet will be discarded. Buffering is good for jitter as well. For example, if the minimum end-to-end delay seen by any packet is 1 ms and the maximum is 6 ms, then the delay jitter is 5 ms. As long as the time delay buffer delays incoming packets by at least 5 ms, then the output of the buffer will include all incoming packets properly and timely sequenced.

Silence generation. Currently, most Internet telephony applications simply generate periods of silence to compensate for lost packets and reproduce delayed packets.

This results in the clipping of speech and a loss of its intelligibility when packets are lost in the network, and a distortion of speech when delayed packets are used to reproduce speech.

Voice reconstruction. Voice reconstruction can occur by the receiver attempting to reconstruct the missing segments of speech from correctly received packets preceding the packet or from packets that are lost or delayed. This can be done by the repetition of the last correctly received speech waveform or via the interpolation process.

However the previous methods are not preventive. There exists also another method, which assists in the prevention of delays and losses. It is achieved by having the applications reserve resources (queue space, outgoing capacity etc.) in order to meet a given quality of service. This preallocation of resources applies in both point-to-point and multicast connections. In the case of IP networks a connectionless approach has been chosen. In this case the endpoints provide the information to the routers by periodically sending messages about the status of the connection. If a new route becomes preferred for a given flow, the endpoints provide the reservation to the new

routers on the route. The protocol that has been developed for performing resource reservation is the RSVP (Resource ReSerVation Protocol) and although it does not actually handle packet delays and losses it ensures better QoS.

C. REQUIREMENTS FOR STANDARDIZATION

Before proceeding to the next chapter it would be beneficial to summarize the necessary characteristics of a standard for Internet Telephony.

Independence and neutrality: A standard should be able to be applied on TCP/IP networks where QoS is under question. However the standard itself should be independent of any network topology and should be applicable on both reliable and unreliable networks.

Complexity and Extensibility: An ideal standard should not add large amount of overhead to the system that it is going to be applied, therefore complexity should be kept at a bare minimum. On the other hand it should allow space for further improvements and enhancements.

Address format flexibility: All possible addressing schemes should be allowed and clearly defined. These include IP-addresses, e-mail, aliases, telephone numbers etc.

Signaling and Call-Control: These include connection establishment procedures and message exchange. At the end-user level all should be transparent. The connection should resemble a standard telephone connection with all the additional functionality (call forward, call transfer, ad-hoc etc.).

Multicast Capabilities: These should be included without adding further complexity.

Security: A very important factor. The standard should define authentication and encryption techniques to be used.

Protocol encoding: The choice of the correct language for implementation is also a serious matter. Text based formats (i.e., HTTP) are flexible and quite easy but they are rather slow. High level languages may be faster but more complex. The choice is up to the working groups.

Internationalization: Whatever the encoding, the types of messages that will be exchanged should be in a format universally recognizable.

Adoption of existing standards: Clearly a new standard should adopt existing specifications derived from previous standards and protocols. Modularity is the major issue here. For example bandwidth reservation requirements through the Internet should be left to be handled by RSVP

since this is the proper mechanism that can accomplish this. Speech compression algorithms already exist and neatly specified in standards as it was previously shown.

At a very minimum those were the most important matters that a standard should be able to address, define and resolve. Nevertheless, the most serious problems of Internet Telephony, packet delay and packet losses were left for last. We will see how the H.323 takes advantage of RSVP's unique features in order to provide soft support for QoS. Also we will see how the H.323 performs with all the given tasks listed above.

III. OVERVIEW AND ANALYSIS OF H.323

In this chapter we will present the basic features and major components of H.323. The scope and architecture of the standard will be explained and analyzed. The standard's features for QoS in IP networks will also be presented. Finally, the operation of the standard will be explained with the presentation of two call scenarios (with and without Gatekeeper). The chapter ends with a synopsis on the H.323 characteristics.

A. SCOPE OF H.323

H.323 covers the technical requirements for multimedia communications systems in those situations where the underlying transport is a Packet Based Network (PBN) that may not provide a guaranteed Quality of Service (QoS). These networks dominate today's corporate desktops and include packet-switched TCP/IP and IPX over Ethernet, Fast Ethernet and Token Ring network technologies. The H.323 is basically a collection of standards and it is based heavily on the ITU multimedia protocols, including H.320 for ISDN, H.321 for B-ISDN and H.324 for PSTN terminals. The recommendation itself describes the components and basic functionality of an H.323 terminal with emphasis in call

signalling, and establishing and termination of conferences. Therefore the standard is limited to the definition of the media compression algorithms, packet format, signalling, and flow control. The work on the standard began on May of 1995. The first version was finished and was approved by the ITU in June 1996. Currently the standard is in its second version (H.323v2) with version 3 on the way.

The scope of H.323 does not include the LAN itself or the transport layer that may be used to connect various LANs. Only elements needed for interaction with the Switched Circuit Network (SCN) are within the scope of H.323. An implementation of an H.323 system is often called an H.323 stack. It is important to emphasize that H.323 is not Network-specific. Moreover it does not contain explicit specifications for IP networks, as it is supposed to lie in a higher level than the transport protocol itself. H.323 uses a binary representation for its messages, based on the Abstract Syntax Notation (ASN.1). The standards that are referenced in the H.323 are shown in Table 4:

Table 4: H.323 collection of standards

Name	Description
H.323	System Document
H.225.0	Describes the H.225 layer were video audio, data and control streams are formatted into messages for output to the network interface and defines the logical framing, sequence numbering, error detection and error correction.
H.245	Describes the system control unit where signaling is provided.
G.7xx	Audio Codecs. These include G.711 (mandatory), G.722, G.728, G.729, G.723.1, and MPEG 1 audio.
Q.931	ISDN user-network interface layer 3 specification for basic call control.
T.120	Series of communication and application protocols and services that provide support for real-time, multipoint data communications
RSVP	Resource ReSerVation Protocol (IETF)
RTP	Real-time Transport Protocol (IETF)
RTCP	Real-time Transport Control Protocol (IETF)

B. COMPONENTS

H.323 introduces four major components for a network-based communications system: Terminals, Gateways, Gatekeepers, and Multipoint Control Units. The role of each component is explained in the paragraphs below.

1. Terminal

An endpoint on the network which provides for real-time two-way communications with another H.323 terminal, Gateway, or MCU. The communication consists of control, indications, audio moving color video pictures, and/or data between the two terminals. Two versions can be implemented:

- Corporate Network (high quality)
- Internet (optimized for low bandwidth 28.8/33.6 - G.723.1 and H.263)

A terminal has also a built in Multipoint capability for *Ad Hoc* conferences.

2. Gateway

An endpoint on the network, which provides for real-time, two-way communications between H.323 Terminals on the packet based network and other ITU Terminals on a switched circuit network, or to another H.323 Gateway. Basically a Gateway is not required if connections to other networks are not needed, since endpoints may directly communicate with other endpoints on the same LAN.

A GW is an optional element in an H.323 conference and it performs the following functions:

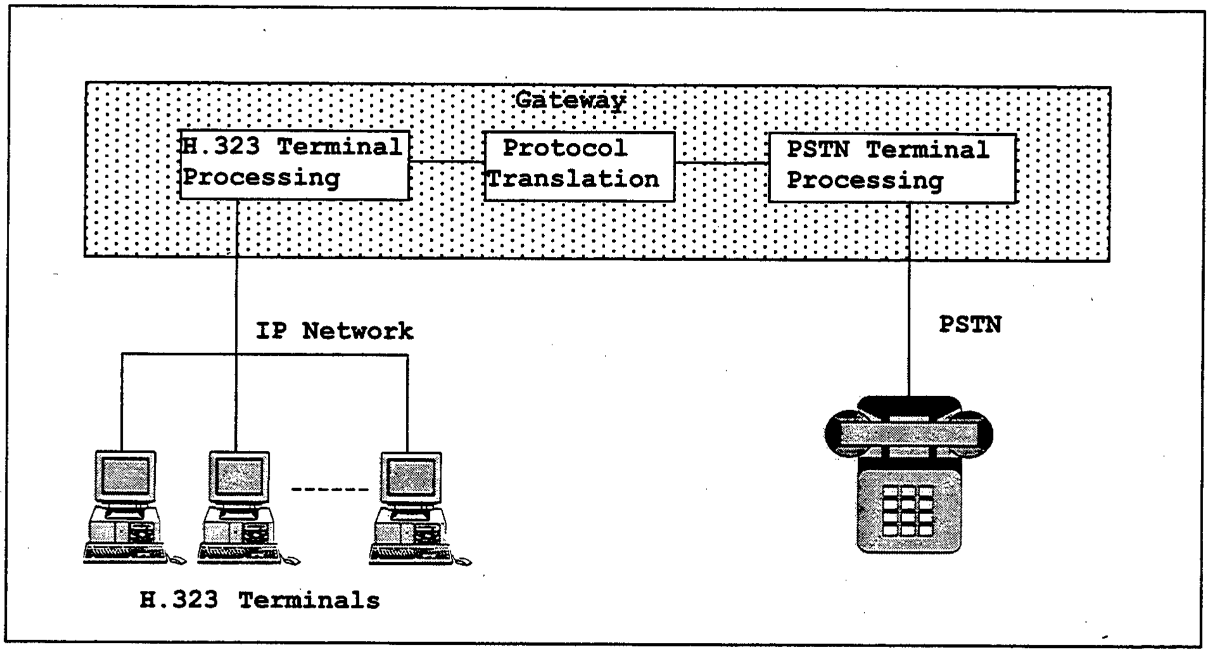


Figure 6: H.323 Gateway

- Provides worldwide connectivity and interoperability from LAN to H.320, H.324, regular Plain Old Telephony System (POTS) telephones.
- Maps Call Signaling (Q.931 to H.225.0).
- Maps Control (H.242/H.243 to H.245).
- Performs Media Mapping (FEC, multiplex, rate matching, audio transcoding, T.123 translation)

3. Gatekeeper

The Gatekeeper (GK) is another optional H.323 entity on the network that provides address translation and controls access to the network for H.323 terminals,

Gateways and MCUs. The GK is a very important component The GK functions include the following:

- Address Translation
  - H.323 Alias to transport (IP) address based on terminal registration
  - "email-like" names possible.
  - "phone number like" names possible.
- Admissions control.
  - Permission to complete call.
  - Can apply bandwidth limits.
  - Method to control LAN traffic.
- Management of gateway.
  - H.320, H.324, POTS, etc.
- Call Signaling.
  - May route calls in order to provide supplementary services or to provide Multipoint Controller functionality.
- Call Management, Reporting and Logging.

The GK is an important component in an H.323 implementation mainly because it can perform address

translations. The endpoints must communicate through the GK using RAS messages as it will be shown below.

4. Multipoint Control Unit (MCU)

The Multipoint Control Unit (MCU) supports conferences between three or more endpoints. Under H.323, an MCU consists Multipoint Controller (MC), which is required, and zero or more Multipoint Processors (MP). The MC handles H.245 negotiations between all terminals to determine common capabilities for audio and video processing. The MC also controls conference resources by determining which, if any, of the audio and video streams will be multicast. The MC does not deal directly with any of the media streams. This is left to the MP, which mixes, switches, and processes audio, video, and/or data bits. MC and MP capabilities can exist in a dedicated component or be part of other H.323 components.

C. ARCHITECTURE

As it was previously mentioned H.323 is a collection of other standards. The layout of the standard's architecture is shown in figure 7. The System Control Unit provides signaling and flow-control for proper operation of a H.323 terminal. H.245 is the media control protocol that

allows capability exchange, channel negotiation, switching of media modes, and other miscellaneous commands and indications. Capabilities exchange is a process using the communicating terminals' exchange messages to provide their transmit and receive capabilities to the peer endpoint. Transmit capabilities describe the terminal's ability to transmit media streams. Receive capabilities describe a terminal's ability to receive and process incoming media

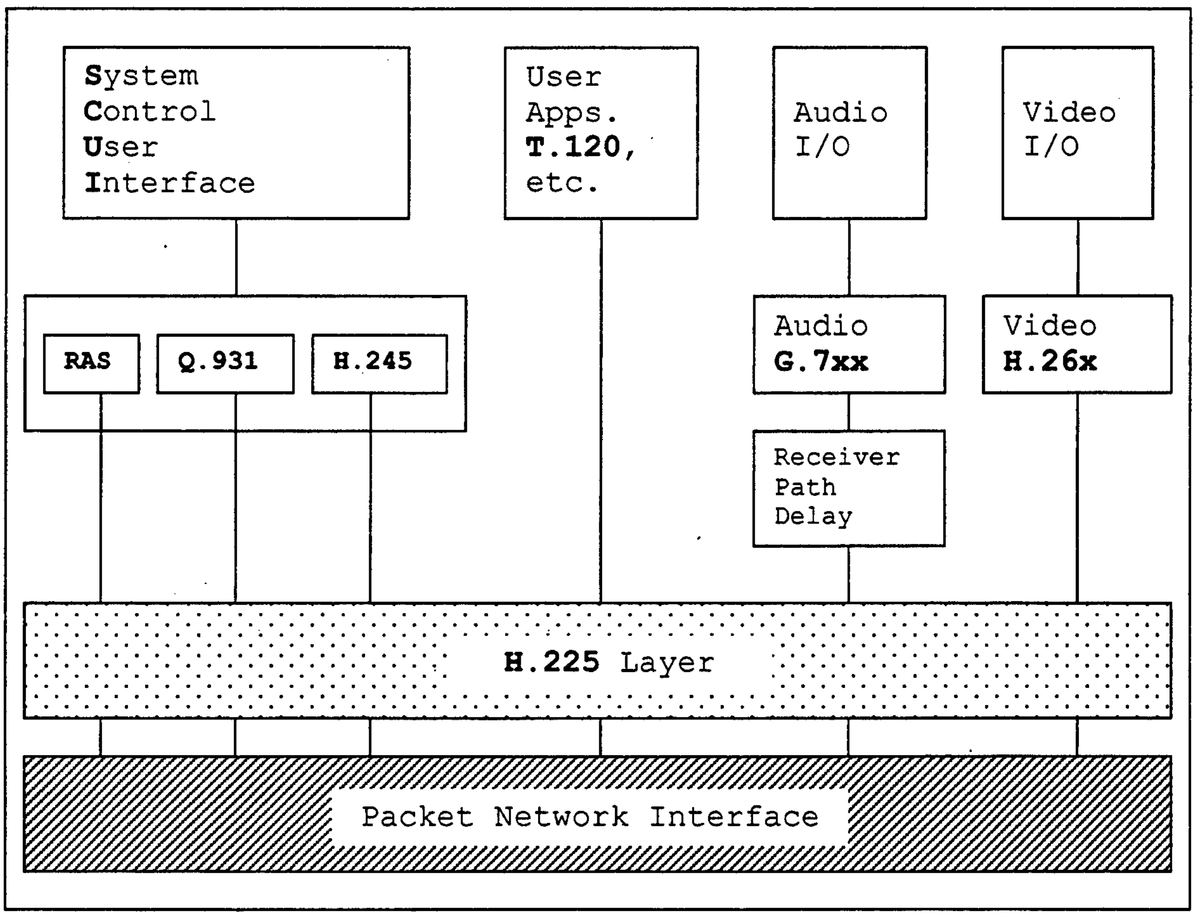


Figure 7: H.323 Architecture

streams.

The Connection Establishment Protocol is derived from the Q.931 specification. The H.225 layer is then responsible for the Call-Signaling Protocol used for admission control terminals and for the formatting of the transmitted video, audio, and data. This procedure is assisted by the Real Time Transport protocol (RTP) and the Real Time Transport Control protocol (RTCP). The RTP performs logical framing, sequence numbering, timestamping, payload distinction, source identification, and occasionally, error detection and correction as appropriate to each media type. The RTCP provides reporting and status indication that may be used by senders and/or receivers to correlate performance on the media streams.

The User Applications that are supported include on-line chat, fax, electronic whiteboards file exchange etc. The T.120 standard is also included, and it is responsible for real-time audiographics conferencing.

All the major ITU-T codecs are supported. G.711 is the mandatory codec for an H.323 terminal. However a terminal may be capable of optionally encoding and decoding speech using other codecs. Since G.711 is a high-bitrate codec (64 Kbps or 56 Kbps), it is not suitable for low-bitrate links.

G.723.1 is the preferred codec and it is the one being adopted by most hardware and software developers. As it can be seen in the diagram H.323 provides definitions for video conferencing as well. The codecs used are described by the H.261 and H.263 recommendations.

D. OPERATION

The core of the standard is dedicated in the detailed description of connection establishment procedures and message exchange using all the possible combinations (i.e., with or without a Gatekeeper). For this purpose the specification is divided in five (5) phases which describe in detail the procedures for calls as shown in Table 3.

Table 5: Call Phases

Phase A	Call setup
Phase B	Initial communication and capability exchange
Phase C	Establishment of audio-visual communication
Phase D	Call Services
Phase E	Call termination

The following paragraphs describe in detail an example of such a procedure between two H.323 endpoints without a gatekeeper. Before proceeding we should clarify how the

different messages are shown. This Recommendation describes the use of three different message types: H.245, RAS and Q.931. To distinguish between the different message types, the following convention is followed. H.245 message and parameter names consist of multiple concatenated words highlighted in bold typeface (**maximumDelayJitter**). RAS message names are represented by three letter abbreviations (ARQ). Q.931 message names consist of one or two words with the first letters capitalized (Call Proceeding).

1. Basic H.323 call without Gatekeeper

The following figure displays all the necessary messages exchanged between two endpoints. The messages have sequence numbers that correspond to the origins of the arrows. Endpoint-1 initiates the call by sending a "Setup" message (1) to Endpoint-2 containing the destination address. Endpoint-2 responds by sending a "Alerting" message (2), followed by a "Connect" message (3) if the call is accepted. This completes the call establishment signaling phase. The next phase, which involves the exchange of H.245 messages then, begins. Both endpoints exchange their terminal capabilities with "**terminalCapabilitySet**" (4). Also they proceed and acknowledge each other's message with

"**terminalCapabilitySetAck**" (5). The next bunch of messages (6-8) is called Master/Slave determination and it is used to resolve conflicts between two endpoints that can both be the MC for a conference, or between two endpoints that are

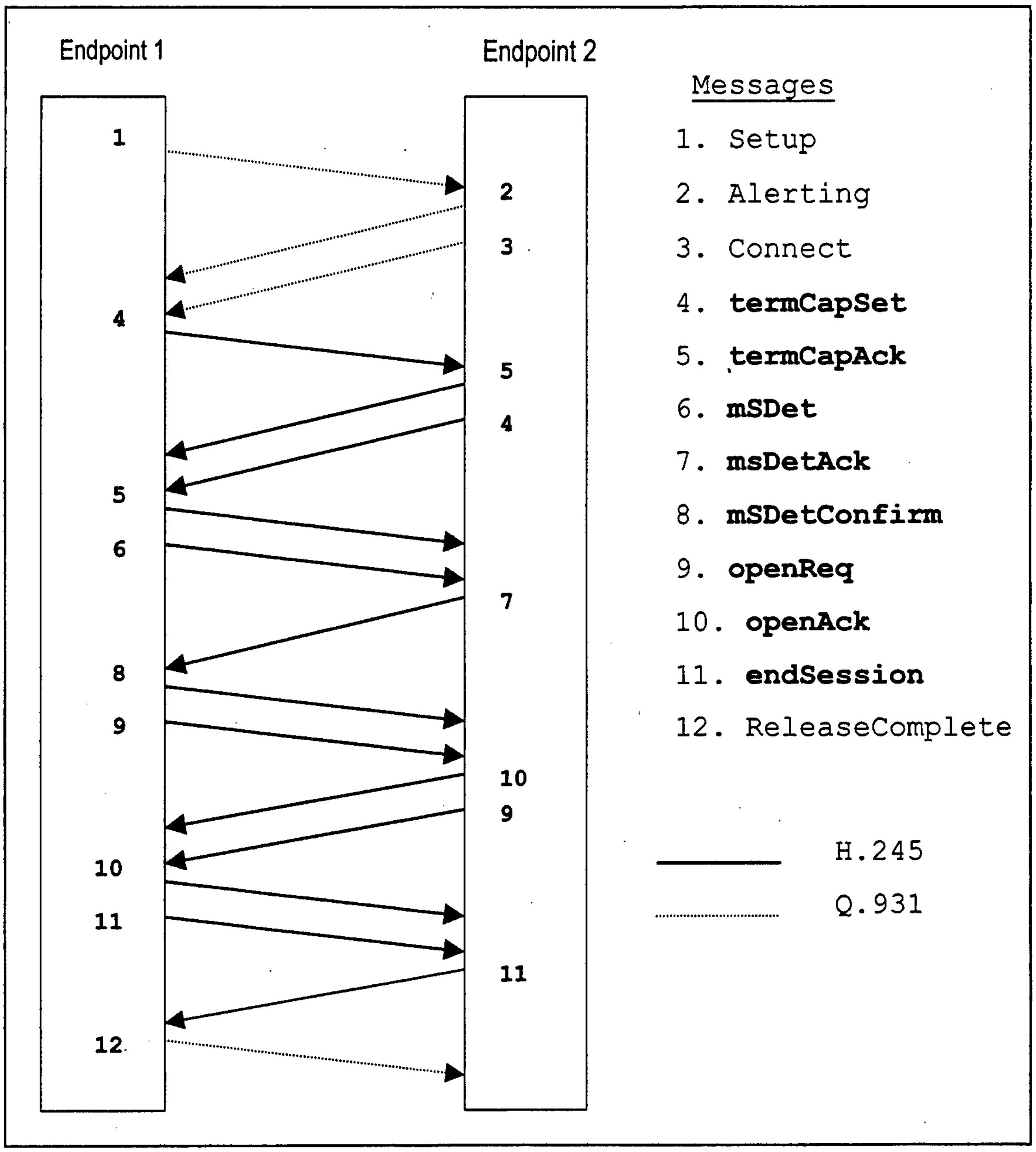


Figure 8: Messages Exchanged between two Endpoints (Without Gatekeepers)

attempting to open bi-directional channels at the same time. During this procedure the two endpoints exchange random numbers in the H.245 "**masterSlaveDetermination**" message to determine the master and slave endpoints. After the two the endpoints proceed to the opening of logical channels (9-10). Unidirectional channels are opened for audio and video while bi-directional ones are opened for data. At this point several other messages may be exchanged, concerning change of media format, change of bitrate, etc. The call is terminated when one of the endpoints sends an "**endSession**" message. In the example of figure 5 Endpoint-1 sends the "**endSession**" (11), Endpoint-2 responds with the same message and the call ends with Endpoint-1 sending a Q.931 "ReleaseComplete" (12) message.

2. Basic H.323 call with a Gatekeeper

In this case the endpoints attempt to find and register with a Gatekeeper before they begin their exchange of messages as shown in figure 8. The endpoints may register with the same or with different gatekeepers. In order to achieve this both endpoints multicast the GatekeeperDiscovery (GRQ) request. The GK may reply either with a GatekeeperConfirm (GCF) or a GatekeeperReject (GRJ) message. The endpoints then register their names with the

RegistrationRequest (RRQ) message and the gatekeeper acknowledges with RegistrationConfirm (RCF) or denies with a RegistrationReject (RRJ) message. This registration allows the endpoints to make the call using user-friendly aliases (e-mails, telephone numbers etc.) instead of the transport address.

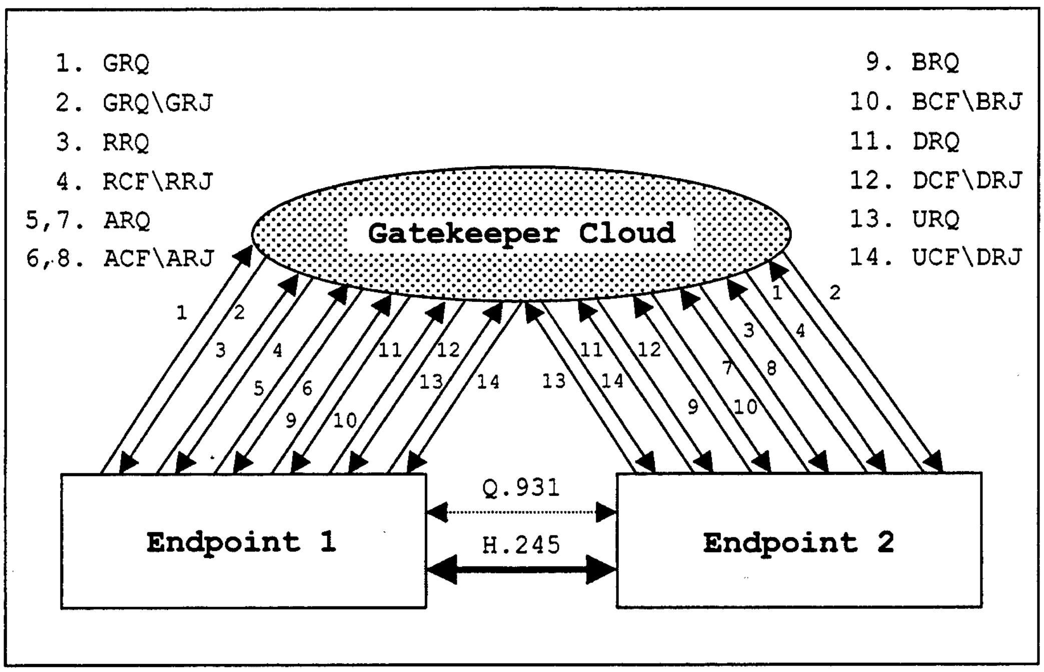


Figure 9: H.323 call with Gatekeeper

The users may then initiate their calls through the endpoints by requesting admission from the gatekeeper using an AdmissionRequest (ARQ) message. The GK accepts or denies the request with AdmissionConfirm (ACF) or AdmissionReject (ARJ). After that the endpoints may start exchanging Q.931 and H.245 messages as shown above. The call ends when both

endpoints send a DisengageRequest (DRQ) message. The GK replies with DisengageConfirm (DCF) or DisengageReject (DRJ). Also the endpoints may unregister with the URQ/UCF/URJ messages. It is worth noticing that during a H.245 session the endpoints may require more bandwidth from the GK with the BRQ/BCF/BRJ messages.

3. QoS Support for H.323

H.323 recommends the use of transport level resource reservation mechanisms to fulfill the QoS requirements of real-time video and audio streams. Although the transport level resource reservation mechanisms themselves are beyond the scope of H.323, the standard contains an appendix that provides some definitions to prevent conflicting interoperability issues.

RSVP is the transport level signalling protocol for reserving resources in unreliable IP-based networks. Using RSVP, H.323 endpoints can reserve resources for a given real-time traffic stream based on its QOS requirements. If the network fails to reserve the required resources, or in the absence of RSVP, only best-effort delivery of the packets is possible.

When an endpoint requests admission with a Gatekeeper, it should indicate in the ARQ message whether or not it is

capable of reserving resources. The Gatekeeper should then decide, based on the information it receives from the endpoint and on information it has about the state of the network, either:

- to permit the endpoint to apply its own reservation mechanism for its H.323 session; or
- to perform resource reservation on behalf of the endpoint; or
- that no resource reservation is needed at all. Best-effort is sufficient.

The specific field in H.225 RAS signaling to permit this functionality is the **TransportQOS** field. In addition to **TransportQOS**, an endpoint may also calculate and report the bandwidth it currently intends to use in all channels of the call. This bandwidth may be reported in the **bandWidth** field of the ARQ message independent of the decision by the endpoint to use RSVP signaling or not. In addition, if bandwidth requirements change during the course of the call, an endpoint may report changes in bandwidth requirements directly to the Gatekeeper.

RSVP reservations can only be made by network entities, which are in the path of media flow between

endpoints. It is possible through Gatekeeper routed call signaling to route media streams through a Gatekeeper. However, most of the time media channels will be routed between endpoints without passing through the Gatekeeper. If a Gatekeeper decides to route media streams, then the procedures followed should be identical to those for RSVP signalling directly from the endpoints.

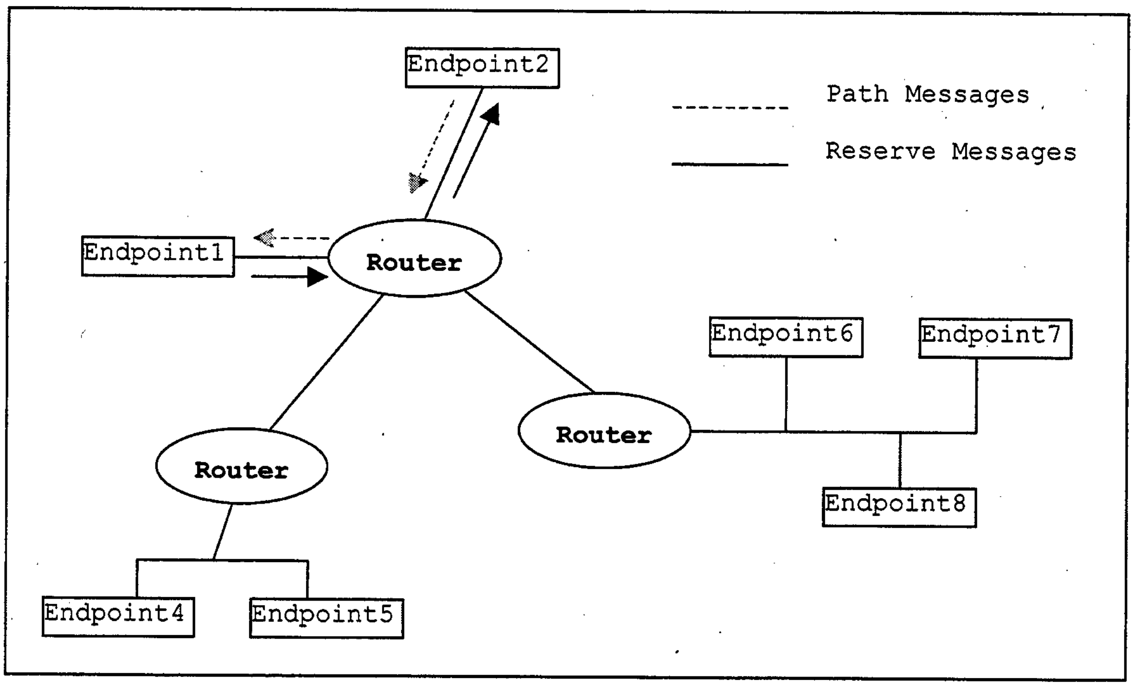


Figure 10: Resource Reservation for a Point-to-Point Connection

In Figure 10, Endpoint1 wishes to send a media stream to Endpoint2. In other words, it has to open a logical channel to B. RSVP signaling for resource reservation may be a part of the opening logical channel procedure. Endpoint1 may cause RSVP *Path* messages to be sent out to 2.

These *Path* messages go through routers and leave "state" on their way tracing towards 2. *Path* messages contain the complete source and destination addresses of the stream and a characterization of the traffic that the source will send. Endpoint2 now can use the information from the *Path* to make the RSVP *Resv* request for the full length of the path. *Resv* messages contain the actual reservation and will generally be the same as the traffic specification in the *Path* message.

RSVP is only a signaling protocol. Together with the appropriate QoS services (e.g., guaranteed QoS or controlled-load service), scheduling mechanisms (e.g., weighted fair queuing), and policy-based admission control module (e.g., local policy manager), RSVP is capable of satisfying the QoS requirements of H.323 conference participants. In addition, RSVP is designed for point-to-point links. If a path traverses a shared link, RSVP invokes the appropriate resource reservation mechanism for the specific shared medium. All the mechanisms mentioned in this paragraph are controlled completely from within RSVP. Therefore, all that an H.323 endpoint needs is RSVP signaling.

4. H.323 features for IP Telephony

The initial environment, for which H.323 was designed, was the corporate network environment, primarily LANs. Wide Area Network (WAN) access was to be gained by using gateways to H.320JISDN. During the implementation of Revision 1 of H.323, it became apparent that IP telephony was gaining popularity and relevance as various infrastructure elements were improved upon. A number of proprietary IP-based telephones were creating many small islands that could not communicate with one another. Recommendation H.323 was supposed to provide a good basis for establishing a universal IP voice and multimedia communication in larger, connected networks. With Revision 2 of the Recommendation, new additions and further extensions were added specifically to make it more suitable for IP telephony.

By using Q931 as its basis for establishing a connection, H.323 allows for relatively easy bridging to the public switched telephone networks (PSTN) and circuit-based phones. The required voice codec of G.711 also allows for easy connections to the legacy networks of telephones. The uncompressed 64kb/sec stream can be translated between digital and analog media. One of the addressing formats

provided in Recommendation H.323 is the E.164 address. This is another ITU Recommendation that specifies standard telephone numbers (e.g., the digits 0-9, * and **#**). These addresses, which ultimately map onto the IP addresses for the H.323 endpoints, allow regular telephones to "dial" them. Gatekeepers provide the final important element for IP telephony. Gatekeepers supply the ability to have integrated directory and routing functions within the course of the call setup. These operations are important for real-time voice or video when resources must be balanced, and points of connectivity are highly dynamic. The gatekeeper functions, which provide call permission and bandwidth control, enable load monitoring, provisioning, and ultimately, commercial-grade IP telephony service.

However it has to be noted that the specification does not contain any explicit directions for IP networks while it contains an entire appendix for telephony over ATM. One would expect a more detailed reference on IPv4 and its successor IPv6.

E. SYNOPSIS

Recommendation H.323 describes the procedures for point-to-point and multipoint audio and video conferencing

over packet-switched networks. In addition to video conferencing terminals, H.323 describes other H.323 entities including gateways, gatekeepers, and MCUs. Gateways allow interoperation of H.323 systems with other audio/video conferencing systems on ISDN, PSTN, and other transports. Gatekeepers provide admission control and address translation to H.323 endpoints. H.323 makes use of RTP, RTCP and allows for resource preallocation with the combined use of RSVP. The core of the recommendation describes the procedures for call establishment and the messages exchanged between endpoints.

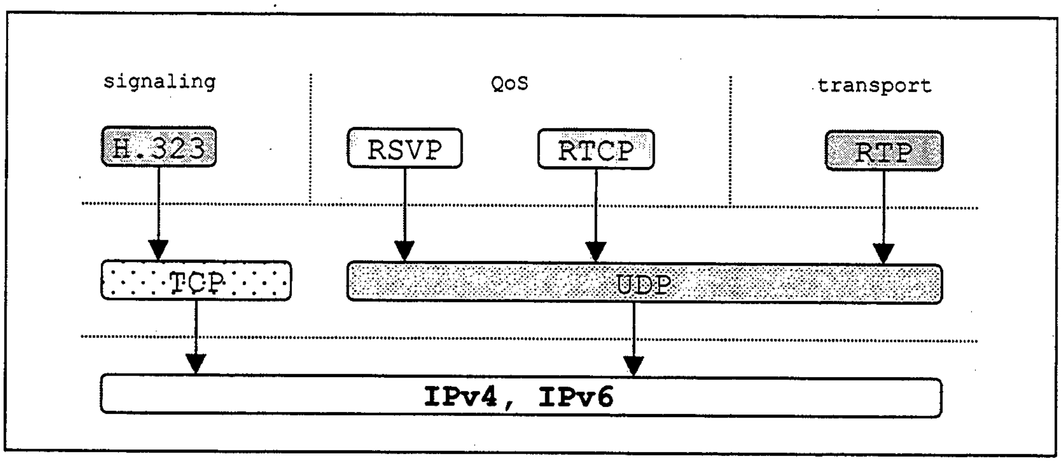


Figure 11: The Internet Telephony Protocol Stack

Figure 11 shows with more clarity what is the role of H.323 in the Internet Telephony Protocol Stack. The first impressions for H.323 are that it does not attempt to re-invent the wheel but it adopts the features of other

protocols instead. However it seems somewhat complex. The following chapter outlines the advantages and disadvantages of the H.323 standard along with a short comparison with its counterpart, the IETF's SIP.

IV. EVALUATION OF H.323

In this chapter we will proceed in the evaluation of H.323 with regards to its complexity, extensibility, interoperability, modularity and other features. A simple call-setup example is also presented in order to display real-life performance considerations on H.323. The chapter ends with a comparison between H.323 and SIP.

A. COMPLEXITY

It was mentioned earlier that an ideal standard should keep the level of complexity at a bare minimum. H.323 is a rather complex standard. The specification itself is 124 pages. If we add the specifications of the other basic standards that are referenced (not including ASN.1) then the sum total comes to approximately 740 pages.

H.323 uses a binary representation for its messages based on ASN.1 and the packed encoding rules (PER). ASN.1 generally requires special code-generators to parse, which in turn create code difficult to debug and maintain. ASN.1 allows any data to be represented in an unambiguous, textual form that is used as a template or schema. After looking over the syntax of ASN.1 one might consider it to be not only complex, but also complicated. This is because

it was designed to be machine-independent. However ASN.1 is also considered flexible and this is a reason why someone could debate over its suitability for H.323. It only remains for us to see the approach of other standards (i.e., SIP) on this matter for a more concrete opinion.

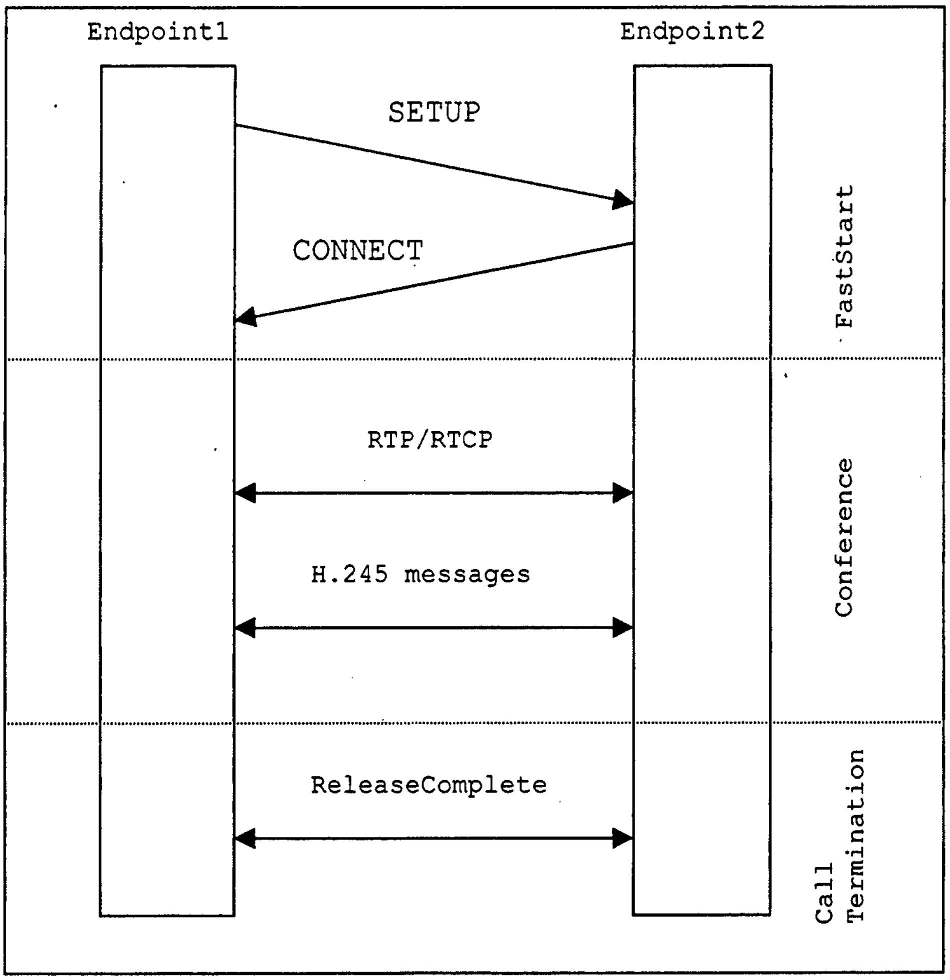


Figure 12: FastStart Procedure

The use of several protocol components adds more complexity to the standard. In the call establishment with a Gatekeeper for example all messages (Q.931, H.245, and

RAS) mix together, producing a lengthy and complex procedure. This was observed from the early days of the standard. Therefore, in version 2 the concept of "FastStart" was introduced (H.323 spec. 8.1.7).

FastStart establishes bi-directional media in one round trip time of messages (discounting the establishment of the actual TCP connection). The messages that normally occur after H.245 establishment are carried along with the Setup-Connect exchange. This facility allows instant audio connection that resembles the regular phone call model as opposed the lengthy H.323 startup procedures. The weak point though is that instead of fully adopting the FastStart procedure, the standard allows any procedure to be used; this means that the endpoints, gateways and gatekeepers must support all of them.

Finally the use of many protocols leads to duplication of functionality. For example H.323 we saw that H.323 makes use of RTP and RTCP. RTCP has been engineered to provide various feedback and conference control functions. Similar functions are built into H.245. The result is redundancy.

B. EXTENSIBILITY AND INTEROPERABILITY

Internet Telephony is expanding extremely rapidly. H.323 needs to be continuously adjusted in order to be able to cope with the growth. So far it has performed satisfactorily. There have been only three years since the first introduction of the standard and now we are heading for version 3. However, it requires full backward compatibility from each version to the next. As various features are added or removed the size of the encoding will increase.

A major issue that affects both extensibility and interoperability is the audio codecs. At this moment hundreds of codecs exist, most of them proprietary. H.323 does support a variety of codecs as was shown, but these are ITU-T standardized codecs only and as such are not for free. This is a significant barrier for many small implementors, including universities. G.723.1 in particular (the most important), is hard to license and very expensive. The reason for this is that 11 vendors (Lucent technologies included) claim rights to some part of the G.723.1 standard.

One positive factor of extensibility for H.323 is that it contains such mechanisms itself. Actually this is a

feature of ASN.1, which contains parameter fields that allow implementors to add their own extensions. This may enable various vendors to experiment with the standard but may lead to interoperability issues since there are no mechanisms to allow the exchange of extensions between two endpoints.

The issue of interoperability is a major headache for implementors and the standardization consortiums. Ideally when an application claims compliance with a standard, it should be interoperable with other applications that are also compliant. This is not true for H.323 for many reasons. The standard offers many options in its specification. An implementor could decide to implement a subset of these options and yet have an application, which in theory is H.323 compatible. The "FastStart" procedure is a prime example on this procedure. The choice of codecs is another. The problem may arise even with applications that have declared H.323v1 or v2 compliance when v3 is already coming, despite the backward compatibility requirement.

The International Multimedia Technical Consortium (IMTC) has focused largely on the area of interoperability by forcing companies to participate in large-scale interoperability tests. Unfortunately the results are

rather disappointing. For example Vocaltec's Internet Phone and Microsoft's NetMeeting (two leading applications) do not communicate with each other, even if they are both H.323 compatible. Needless to say that interoperability should occur in all forms of Internet Telephony (PC-to-PC, PC-to-Phone, etc.). If the industry is to converge on one ubiquitous, feature rich, find-and-connect protocol, (as H.323 claims to be), then we are about to see a tremendous growth of new H.323 endpoints, without interoperability assurance.

C. MODULARITY

This is definitely a strong point of the H.323's features. Clearly the standard makes use of existing technology. The basic services of an Internet Telephony application are handled by other standards. H.245 for capability exchange, Q.931 for call signaling, RTP and RTCP for transfer and RSVP for resource reservation and preventive QoS technics. This modularity will prove to be critical in the future, but H.323 will have to prove that it can easily discard old standards and protocols and accept new ones. It is certain that new mechanisms will evolve in the future (especially with regards to QoS).

H.323 should have the flexibility to allow interoperation of the new mechanisms with the existing technology.

D. OTHER FEATURES

H.323 was originally conceived for use on a single LAN. In the first version the issue of wide area addressing was not a concern. For this purpose the newest versions introduced the concept of the *zone*.

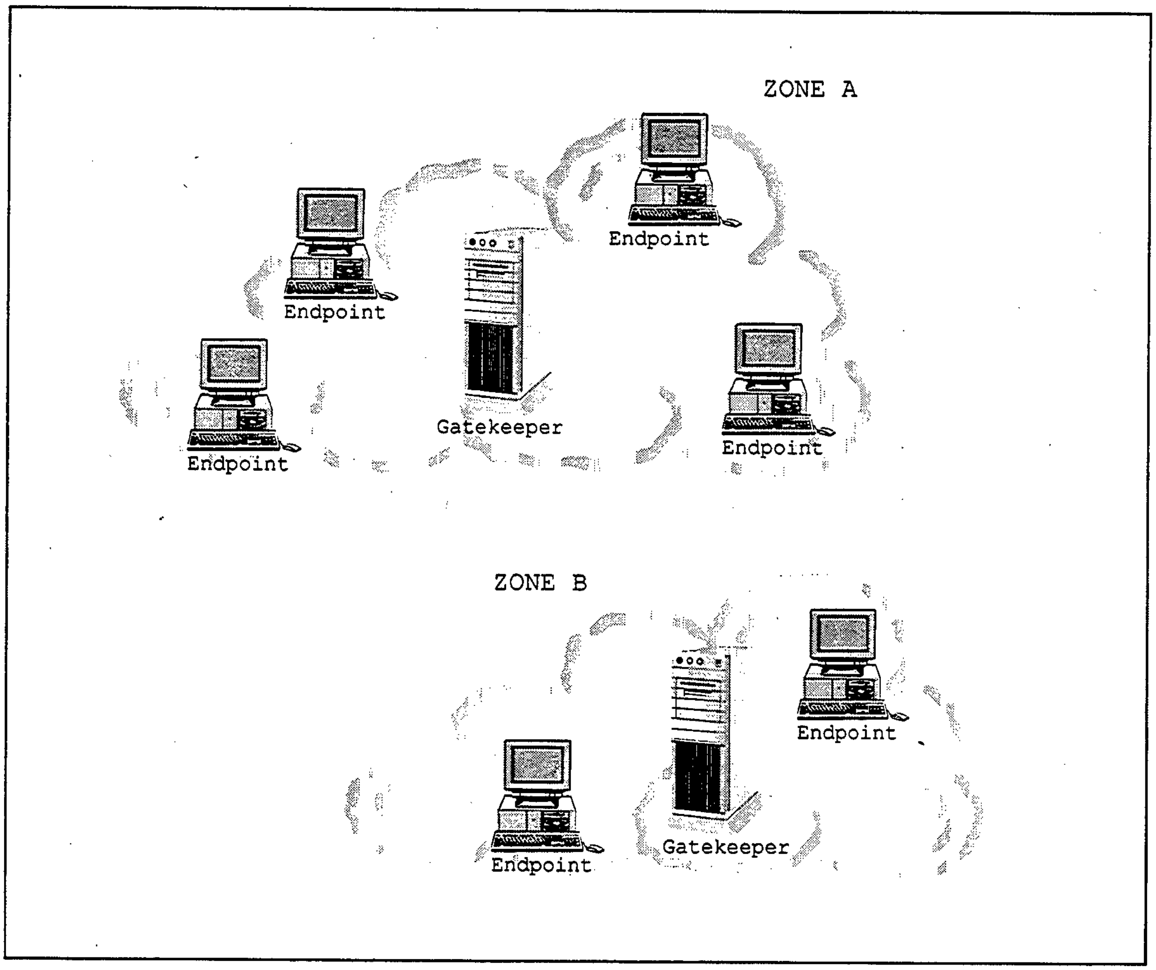


Figure 13: Zone Management

These are simply sets of several H.323 entities grouped together under the supervision and responsibility of a Gatekeeper. Another way to describe a gatekeeper zone is to call it "an administrative domain." Figure 13 shows a zone topology. Consequently the standard defines procedures for user location and GK discovery across zones. This configuration may pose a problem when many entities are to be created in a large environment. The detection of users may become difficult and time consuming.

Another significant feature that may affect the standard's performance in the future is that the gatekeepers during a call are required to remain in a CALL state for the entire duration of a call. This leads to memory consumption and system degradation. The implementors have already begun to realize that. For example Vocaltec's Internet Phone warns the users that H.323 requires additional system resources and should only be used for PC-to-Phone conferences.

The issue of security was left for last. This is considered an optional enhancement for H.323 systems. However if it is going to be provided, it shall be provided in accordance with Recommendation H.235 (H.323 spec. 10.1).

E. H.323 PERFORMANCE CONSIDERATIONS

Depending on whether a gatekeeper is being used or not, establishing an H.323 call can take about a dozen packets and about **6** to **7** round-trip times, including setting up the Q.931 and H.245 TCP connections. For a modem connection, where transmission delays are substantial,

Report for di.uoa.gr [195.134.65.36]

Analysis: Node 'di.uoa.gr' was found in 20 hops (TTL=237). Connections to HTTP port 80 are being rejected.

Ho	Er	IP Address	Node Name	Location	ms	Graph	Network
0		131.120.50.169	slc169.cc.nps.navy.mil	*		0 877	Naval Postgraduate School
1		131.120.50.12	tsb.cc.nps.navy.mil	Monterey, CA 93943	424		Naval Postgraduate School
2		131.120.50.1	-	Monterey, CA 93943	377		Naval Postgraduate School
3	1	131.120.251.2	-	Monterey, CA 93943	350		Naval Postgraduate School
4		131.120.252.101	epsilon.nps.navy.mil	Monterey, CA 93943	309		Naval Postgraduate School
5		131.120.250.2	-	Monterey, CA 93943	320		Naval Postgraduate School
6		135.206.243.10	maewestbr-aip.att-disc.net	San Jose, CA, USA	310		AT&T Government Markets
7		135.206.243.3	nccoscbr-aip.att-disc.net	Oakton, VA 22185	378		AT&T Government Markets
8		138.18.176.1	nccosc-a3-0-1.dren.net	Arlington, VA 22201-4	372		PM Supercomputers Office
9		138.18.175.2	mae-west-a4-0-2.dren.net	San Jose, CA, USA	428		PM Supercomputers Office
10		198.32.136.42	san-jose3.ca.alter.net	San Jose, CA, USA	398		Exchange Point Blocks
11		146.188.144.134	118.ATM3-0.XR2.SJC1.ALTER.NET	San Jose, CA, USA	396		UUNET PIPEX
12		146.188.146.54	192.ATM3-0.TR2.SCL1.ALTER.NET	Santa Clara, CA, USA	384		UUNET PIPEX
13		146.188.137.169	107.ATM6-0.TR2.NYC1.ALTER.NET	New York, NY, USA	408		UUNET PIPEX
14		146.188.179.37	198.ATM7-0.XR2.NYC4.ALTER.NET	New York, NY, USA	473		UUNET PIPEX
15		152.63.21.205	188.ATM7-0.GW9.NYC4.ALTER.NET	New York, NY, USA	467		UUNET Technologies, Inc.
16		157.130.19.22	dante3-gw.customer.ALTER.NET	-	476		UUNET Technologies, Inc.
17		212.1.200.22	athens-ten-155-us.grnet.gr	Athens, Greece	600		DANTE US service - serial li
18		194.177.208.46	uoa-athens.grnet.gr	Athens, Greece	582		Greek Research and Techn
19		195.134.70.11	-	Athens, Greece	543		University of Athens
20		195.134.65.36	di.uoa.gr	Athens, Greece	613		University of Athens

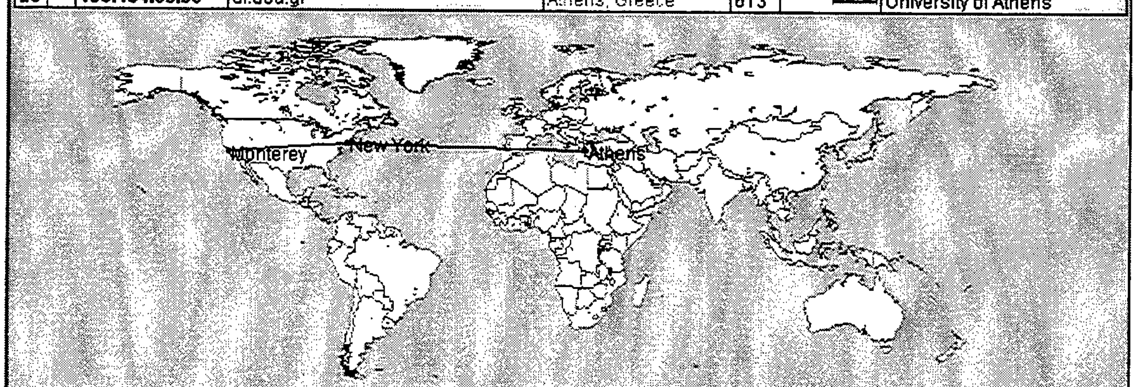


Figure 14: Visual Traceroute

setting up an H.323 call can take several seconds. In order to gain some understanding of the problem the following scenario is introduced.

Consider a call from the Naval Postgraduate School in Monterey (California) to a host in the Department of Informatics in the University of Athens (Greece). Such a call goes almost through half of the earth's perimeter, **6848 miles (11020 km)** to be exact. Geographical distance is not the issue here; it is given for informative purposes. Hop distance is the important factor. However, the two locations were chosen in an attempt to replicate a long distance call. A traceroute measurement was taken using VisualRoute (by Datametrics Systems).

As it can be clearly seen the *traceroute* command resulted in a **613 ms** average Round Trip time (RTT) despite the fact that the expected RTTs are estimated in an average of **300 ms** during peak hours. Our "call" went through a **28.8 Kbps** dial-up connection followed by 20 hops of unknown quality traffic in-between. Consequently if we had attempted an H.323 call setup during that time we would have required at least **3.5** to **4.5 seconds** for the setup. This is a best-case scenario assuming no packet losses during the setup and without considering the standard's

overhead (Gatekeeper discovery time, etc.) Some vendors expect the actual call-setup times to be **20** to **30 seconds,** which is simply unacceptable to the average telephone customer.

This simplistic approach shows only the "tip" of the iceberg of the round-trip delay. It also shows the problem of Internet Telephony in general because such a high RTT would result in long delays and latency during a telephone call. The ITU-T G.114 recommendation limits RTT to **300 ms** or less for telephone traffic. This performance factor is based on many studies and observations; they conclude that longer delays in a telephone-based conversation, gives the impression to the callers that they are using a half-duplex circuit. (Interestingly, other surveys show some people tolerate large RTTs of up to 800 ms. But this tolerant population is in the minority.)

A single example of-course will not suffice for the actual performance measurement of H.323. As of this writing new tools were developed for the actual performance measurement of H.323 entities. Currently however only analytical modeling can be used combined with some measurements similar with the one above.

Call-setup is only a fraction of the H.323 specification but it is the one that can be easily measured. Our example reveals the nature of H.323, which is logically identical with the nature of traditional telephony. Let us not forget that the standard was introduced by ITU. One could easily expect that the ITU researchers would be more focused in Telephony rather than the Internet. This would make sense and it could be accepted if the Internet did not have its unique attributes concerning the transfer of data packets. The telephone network does not have the same problems. First, the path between the talker and listener is fixed during the call set-up. Second, the circuit switches do not queue the traffic. Rather, the voice channels are time division multiplexed into DS0 channels and sent directly from the input interface on the switch to a corresponding DS0 slot on a preconfigured output interface. The delay through the voice switch is minuscule (it is not even a factor), and fixed.

F. H.323 AND SIP

Considering the fact that the H.323 standard is still in the process of being improved and developed it would be

premature and perhaps unfair to characterize the standard inadequate for Internet Telephony due to its complexity. Especially when it has already been recognized and adopted by major Computer and Networking companies like Intel, Microsoft, Lucent Technologies and others. The fact is though that it doesn't seem to be the perfect solution yet for Internet Telephony. Other research groups have already begun to seek for alternatives. One such approach is the Session Initiation Protocol (SIP) by IETF.

The SIP is a signaling protocol that operates with user agents and user agent servers. The main job of the server is to provide for name-to address resolution and user location. For example, when a user makes a call, the user agent sends an SIP message to a server. The user is unaware of this support operation, but will have given its agent an identifier, such as a phone number. The message is sent to a server by the agent, and at this server, the name may be resolved to an IP address, or the server may redirect (proxy) the message to another server.

SIP allows more than one server to contact the user, and these forked messages are sent to multiple servers. The responses are returned to the agent in such a manner that

the agent can make decisions about the best path for the call.

SIP is an attractive alternative support tool for Internet telephony because:

- It can operate as stateless, or stateful. Thus, a stateless implementation provides good scaleability, since the servers do not have to maintain information on the call state once the transaction has been processed. Moreover, the stateless approach is very robust, since the server need not remember anything about a call.
- It uses much of the formats and syntax of HTTP (Hypertext Transfer Protocol), thus providing a convenient way of operating with ongoing browsers.
- The SIP message (the message body) is opaque; it can be of any syntax. Therefore, it can be described in more than one way. As examples, it may be described with the Multipurpose Internet Mail Extension (MIME), or the Extensible Markup Language (XML).
- It identifies a user with a URI (Uniform Resource Identifier), thus providing the user the ability to initiate a call by clicking on a web link.

The following table summarizes the major differences between SIP and H.323 and focuses on H.323's weak points in comparison with SIP.

Table 6: H.323 vs. SIP

	H.323	SIP
Specification volume	700+ (with all the referenced protocols not including ASN.1)	120+ pages
Call-Setup packets required	Up to 12	4
Complexity	Complex, references a large amount of protocols	Less Complex and shorter
Encoding	Q.931 and ASN.1 PER encoding	Text-based similar to HTTP and RTSP
Multicast signaling	NO	YES
Internationalization	Unicode (BMPString within ASN.1) with generally few textual parameters.	Unicode (ISO 10646-1), encoded as UTF-8, for all text strings
Security	References H.235 and it is optional	SSL and HTTP security
Modularity	Makes use of many existing standards	Less modular
Compression Algorithms	ITU-T only	Can work with any codec

From the IETF's perspective SIP is the way to go. Indeed it seems to be a simple and attractive protocol but at the same time it cannot be fully compared with H.323 because it is limited in signaling specifications. Combined with RTP it could create a more robust solution. This

solution is currently under development with the name of Media Gateway Control Protocol (MGCP) which is in the Internet draft stages. SIP will be a part of MGCP. In the future we will see if there will be a dominating protocol or the landscape will become even more confusing with several protocols and standards struggling with each other.

V. INTERNET TELEPHONY SOFTWARE

A. OVERVIEW

In order to gain a clear understanding of the curent status of Internet Telephony in terms of quality interoperability and compliance with existing standards a test study was conducted on some of the curent software PC-to-PC products that exist out there. Specifically the following 10 programs were tested:

- Internet Phone (Vocaltec)
- NetMeeting (Microsoft)
- Intel Video Phone (Intel)
- CU-SEE-ME (White Pine Software, Inc.)
- GatherTalk (The Chinese University of Hong Kong)
- WebPhone (NetSpeak)
- BuddyPhone (Henry Pfluger)
- PGPfone (Pretty Good Privacy, Inc.)
- Speak Freely (John Walker, Brian Wiles)
- IRIS Phone (IRIS Systems)

The tests were conducted on a simple peer-to-peer network with two PCs running Windows 98 and NT 4.0 connected with a 10-Base-T Ethernet running TCP/IP and also

through a dial-up connection of 28.8 Kbps in a maximum hop distance of up to 20 hops. Therefore the tests reflect the assesment of the quality of these products when operating under ideal network conditions (on a small LAN) and under the real problems of the Internet world. The specifications of the programs were also examined for standards and protocols compliance (the focus was on compression algorithms and H.323). Finally the experiments included interoperability testing simply to show if the applications were capable of communicating with others.

The evaluation of the tests with regards to quality is subjective. But the evaluation of voice quality on the telephone network is subjective as well. A scale of 0 to 10 was devised with the expectation that a score between 7 and 8 represented the current «toll quality» exhibited in the telephone network. The evaluation of 10 would reflect the best possible telephone voice conection wherein the analog loop is terminated into a digital system one time only, sent a very short distance, and then converted back to an analog signal.

Finally it has to be noted that not all the existing programs were tested. The focus was on those programs that offered multiplatform versions because the ideal Internet

Telephony software should not be limited in specific Operating systems and platforms.

B. TEST RESULTS

1. Internet Phone

This was the first commercial software developed for Internet Telephony created by Vocaltec. It allows full duplex or half-duplex conversation and comes with a voice activation feature. It claims that it supports H.323 and offers the user the option to configure the program in terms of H.323 compatibility and choice of compression algorithm. However, the program could not communicate with any of the other programs including the H.323 compliant ones. The program also uses its own DLS servers (named Community Browsers by the company) and it could not connect with any other DLS server. The quality of voice was very good during all our tests. The codecs that are used include GSM and G.711 (mandated by H.323), and some proprietary ones that are developed by Vocaltec.

The interface of the program has a "cellular phone" look and feel and it certainly has lent some of its features to other programs that we tested. One useful feature of the program is that it contains a built-in

monitor which monitors the network traffic during the conversation indicating packet losses and delays (shown in figure 12).

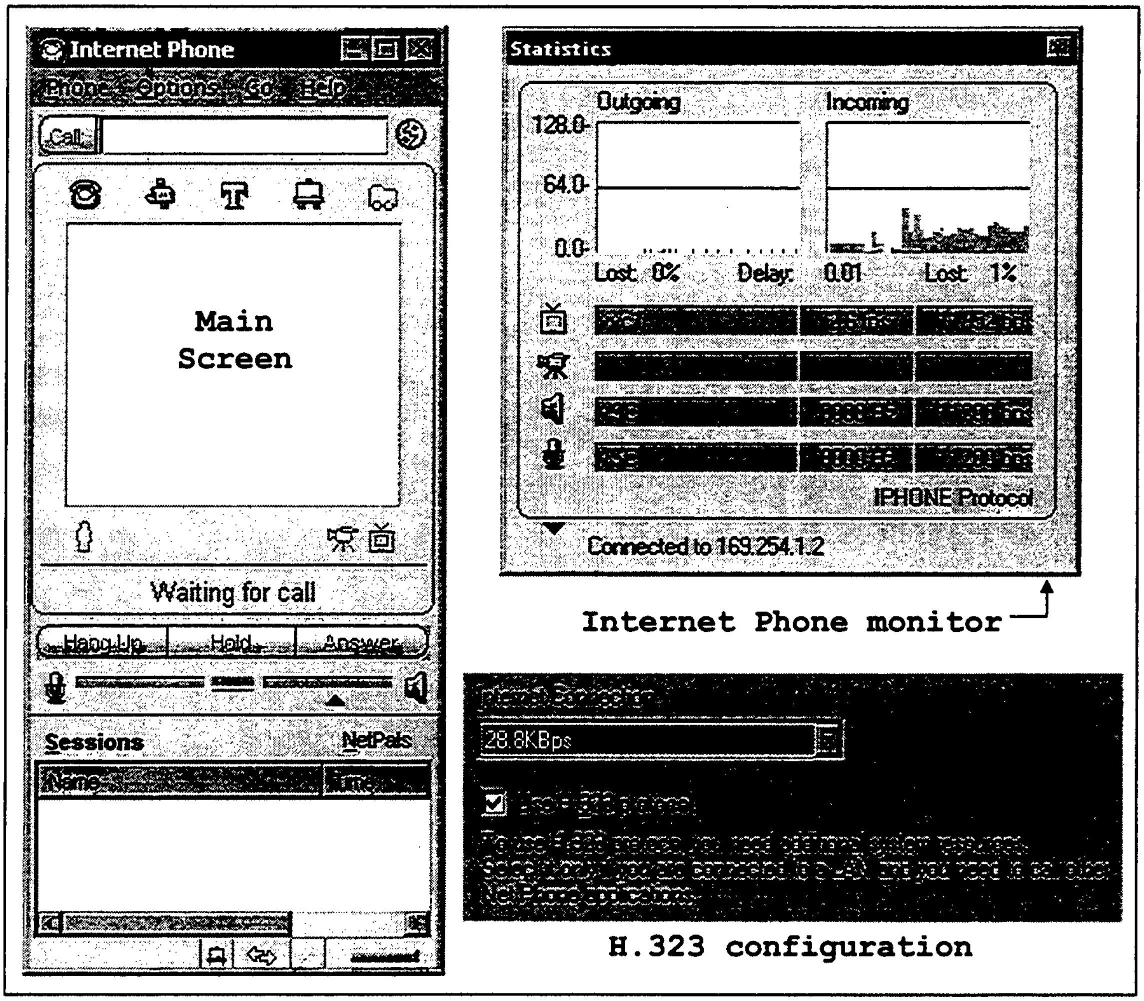


Figure 15: Internet Phone Screenshots

The configuration for H.323 is rather vague. The user can only check or uncheck a checkbox, which allows the use of H.323 standard as shown in Figure 12. Notice the warning

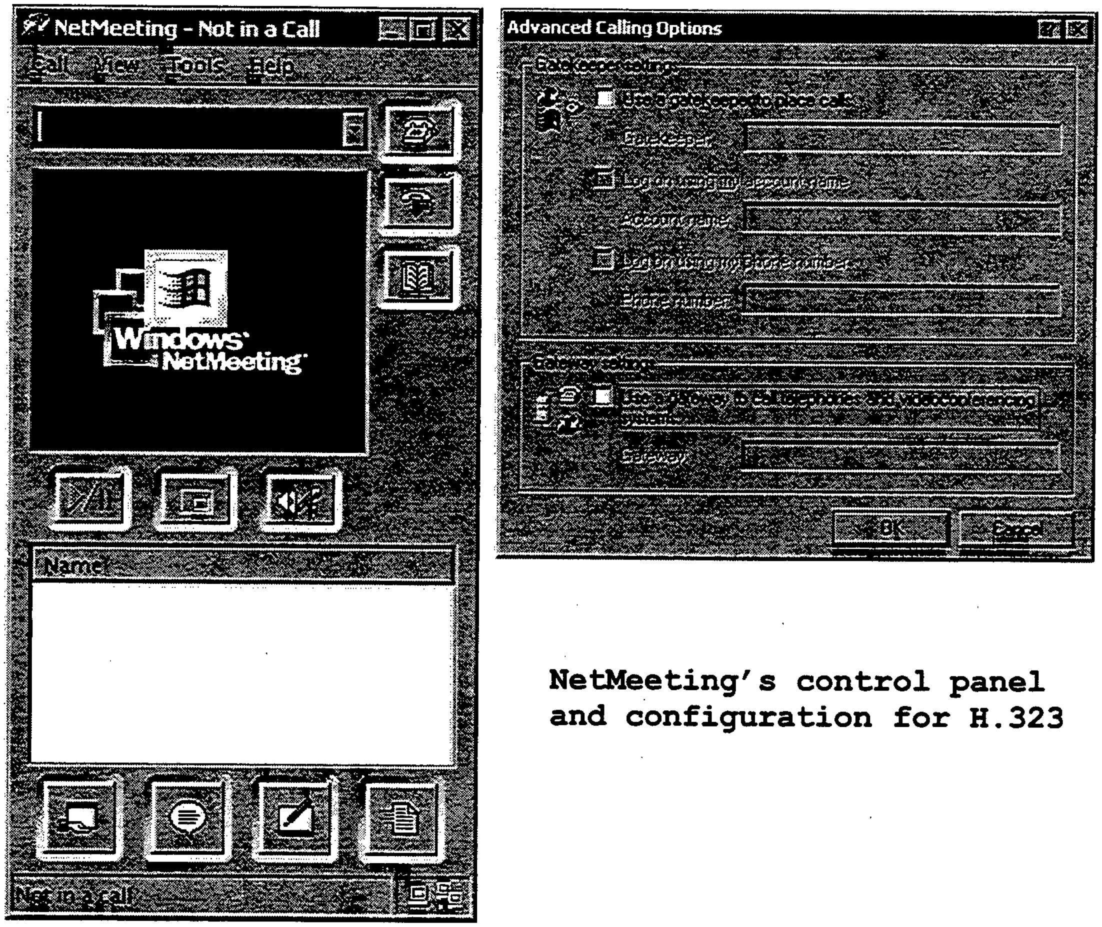
for resources consumption. The results of the program testing are summarized in table 7.

Table 7: Internet Phone Summary of Results

Product name and Version	Internet Phone 5.01
Company or implementor	Vocaltec
LAN quality	8/10
Internet Quality	6/10
H.323 compliance	YES
Audio Compression algorithms	- Vocaltec VSC (8 kHz) - DSP Group TrueSpeech - GSM - Vocaltec VSC (5.5 kHz) - G.711 µ-Law
Interoperability	None
Platforms	- Windows 95/98/NT - MacOS
Address formats	- IP-Address - e-mail - DLS alias
Video support	YES
Other features	- Voice mail - Whiteboard - Text-Chat - File Transfer

2. NetMeeting

This is Microsoft's attempt to gain share on the Internet Telephony battlefield. Microsoft fully supports H.323 and allows the user to choose a Gateway and/or Gatekeeper. The quality of voice was equivalent to that of Internet Phone and one of the best among the software tested. The user can also choose among a list of codecs to use depending on the type of network. The program uses by default G.723.1. Video support is included, as it seems to be the current trend among the telephony programs. The latest version comes with NetMeeting Software Development Kit (SDK) that allows users to create their own applications based on NetMeeting and it provides guidance for embedding the NetMeeting panel in Web Pages. NetMeeting also allows for multi-user conference but the quality of voice is drastically degraded if there are more than two users on-line. NetMeeting is interoperable with other programs but not with Internet Phone even when the same codecs are used. Overall it is a decent product that is freely available with an interface that resembles that of Internet Phone. Currently it can be downloaded with or without Microsoft's Internet Explorer browser.



NetMeeting's control panel and configuration for H.323

Figure 16: NetMeeting Screenshots

NetMeeting supports a variety of codecs including some proprietary ones. Best results are achieved with G.723.1. The program comes with all the major telephony bells and whistles. Calls can be placed either directly using IP-addresses or via the company's own set of DLS servers. All these options are configurable by the user.

Table 8: NetMeeting Summary of Results

Product name and Version	NetMeeting 3.01
Company or implementor	Microsoft
LAN quality	8/10
Internet Quality	5.5/10
H.323 compliance	YES
Audio Compression algorithms	- G.723.1, 8 kHz - Lernout & Hauspie 8 kHz with 8, 12 and 16 Kbps - G.711 A-Law and µ-Law - Microsoft ADPCM 8 kHz
Interoperability	- Intel Video Phone - Webphone - CU-SEE-ME
Platforms	- Windows 95/98/NT - MacOS
Address formats	- IP-Address - e-mail
Video support	YES
Other features	- Voice mail - Whiteboard - Text-Chat - File Transfer - Applications sharing - Remote Desktop sharing

3. Intel Video Phone

Intel's Video Phone was one of the first to use the standards set by the Internet Telephony industry namely the H.323. It offers the same functionality with the previous products but it doesn't allow the user to configure the application manually with regards to standards and codecs.

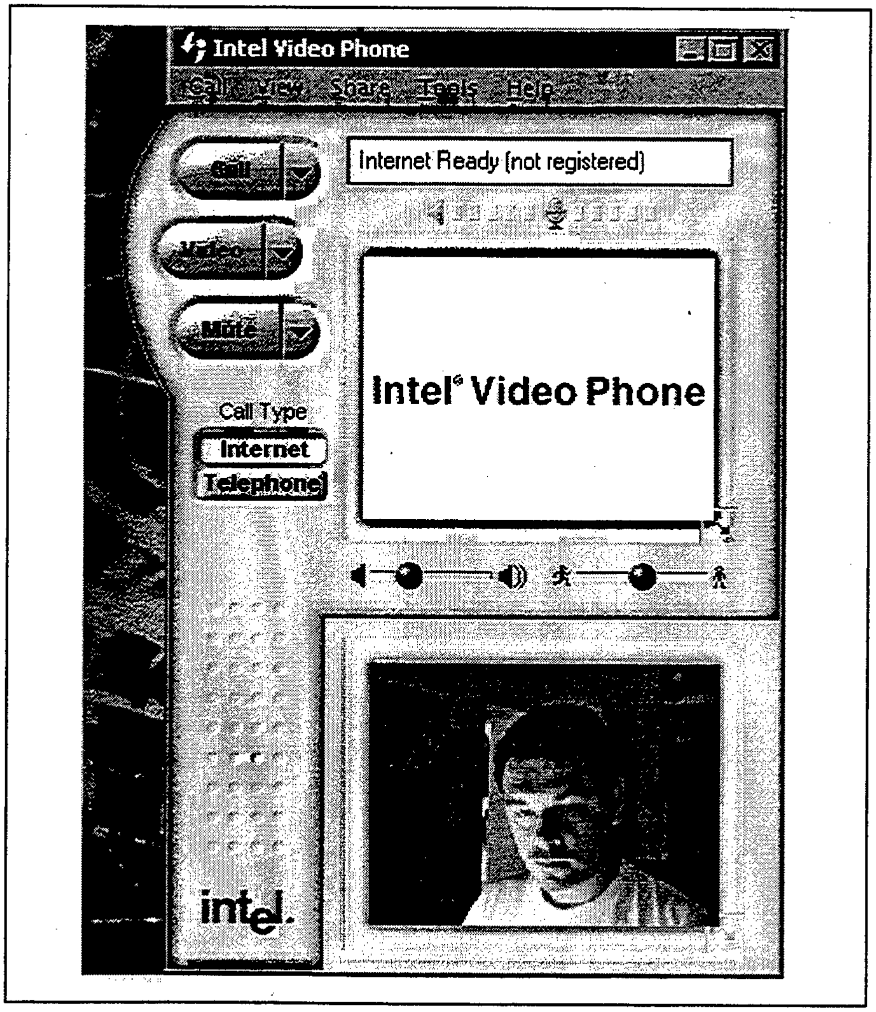


Figure 17: Video Phone's Main Panel

The application claims to be H.323 compliant but there is no configuration available. Also the codecs used are transparent at the user level. The makers of this product have put more weight on simplicity. Automatic wizards do all necessary configuration. The quality of voice was good on the LAN conferences but rather mediocre on the Internet with lots of latencies and losses. Overall very easy to setup and install but with few configuration options.

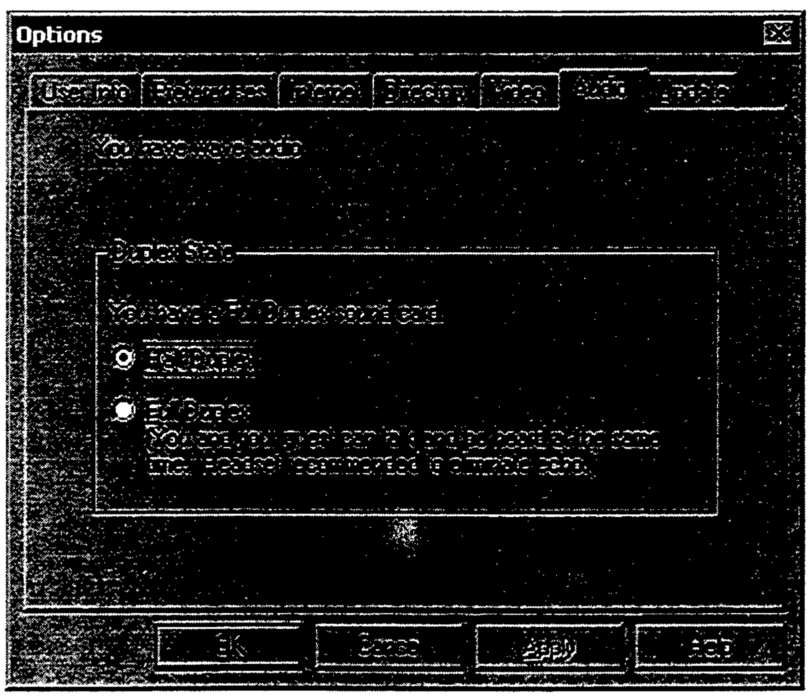


Figure 18: Minimal Audio Configuration

Table 9: Intel Video Phone Summary of Results

Product name and Version	Intel Video Phone 3.1.0.77
Company or implementor	Intel
LAN quality	7/10
Internet Quality	4.5/10
H.323 compliance	YES
Audio Compression algorithms	- Intel Indeo Audio - G.711 - G.723.1
Interoperability	- NetMeting - Webphone - CU-SEE-ME
Platforms	- Windows 95/98
Address formats	- IP-Address - e-mail
Video support	YES
Other features	- Whiteboard - Text-Chat - File Transfer - Applications sharing

The Intel Video Phone uses special flow-control and remote acknowledgment features to change the bandwidth during a video call. These flow-control features help prevent data overrun and allow users to make adjustments

that eliminate extreme audio & video delays that can be caused by sending more data, than the end-to-end Internet connection has bandwidth to carry or than the guest system can decode and display. The current H.323 specification has no flow-control features and there is no provision for changing the data rate during a call. Therefore the flow-control commands used by Intel Video Phone are proprietary and can only be recognized and responded to by other Intel Phones (versions 3.2 & higher). Despite the above the program behaves satisfactorily when used to call other Internet Telephony applications.

4. CU-SEE-ME

This is a another H.323 compatible product with a standard "browser-like" interface. It produced one of the best audio qualities of the tests. The program is highly configurable and allows the user to choose audio codecs directly starting from the well-known low bit-rate G.723.1 up to the company's proprietary codec at 32 kbps. The latter was used for LAN calls while G.723.1 was used for Internet calls. The quality was almost equally good. The difference was a higher latency observed during the Internet calls due to the codec overhead. Although the product claims H.323 compliance it does not support G.711

which is mandatory for H.323 but offers G.723.1 instead for H.323 calls. This is a typical example of the confusion that is caused among the implementors with regards to the right choice of codec. The product can work seamlessly with both NetMeeting and Intel VideoPhone but only through a DLS server. No configuration option is offered for H.323 calls other than the G.723.1 option. The program also allows the user to choose manually the type of Network connection starting from 28.8 kbps modem up to T1 LAN connections.

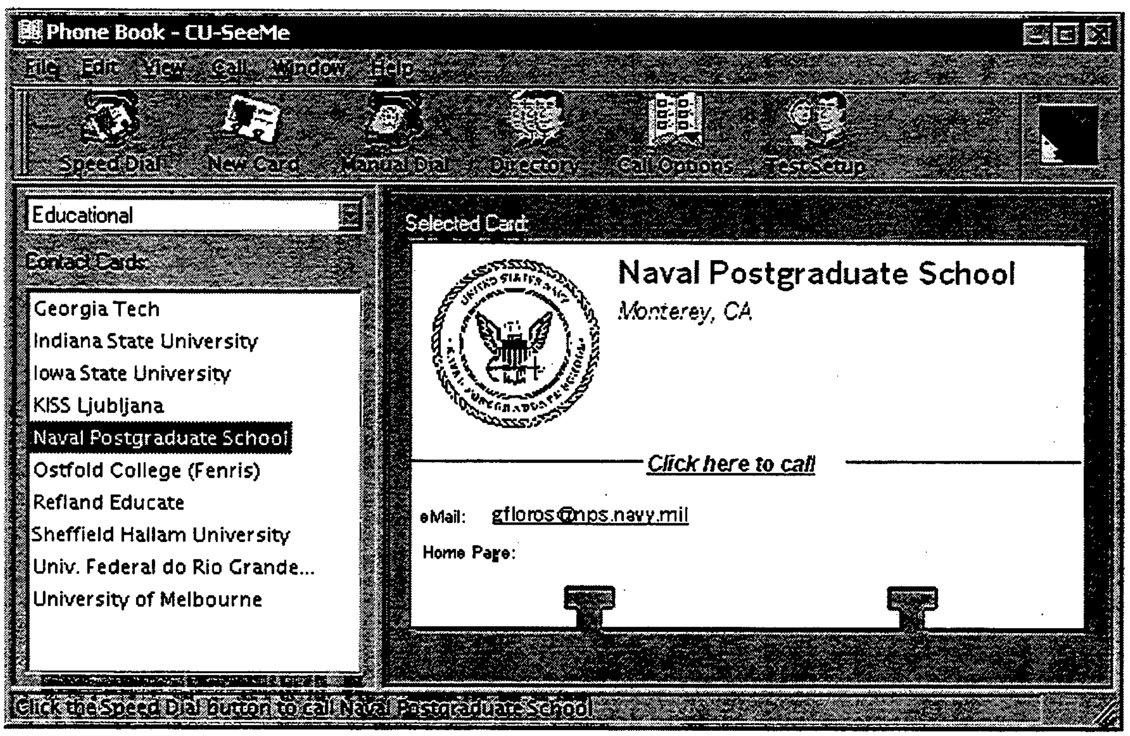


Figure 19: CU-SEE-ME Main Panel

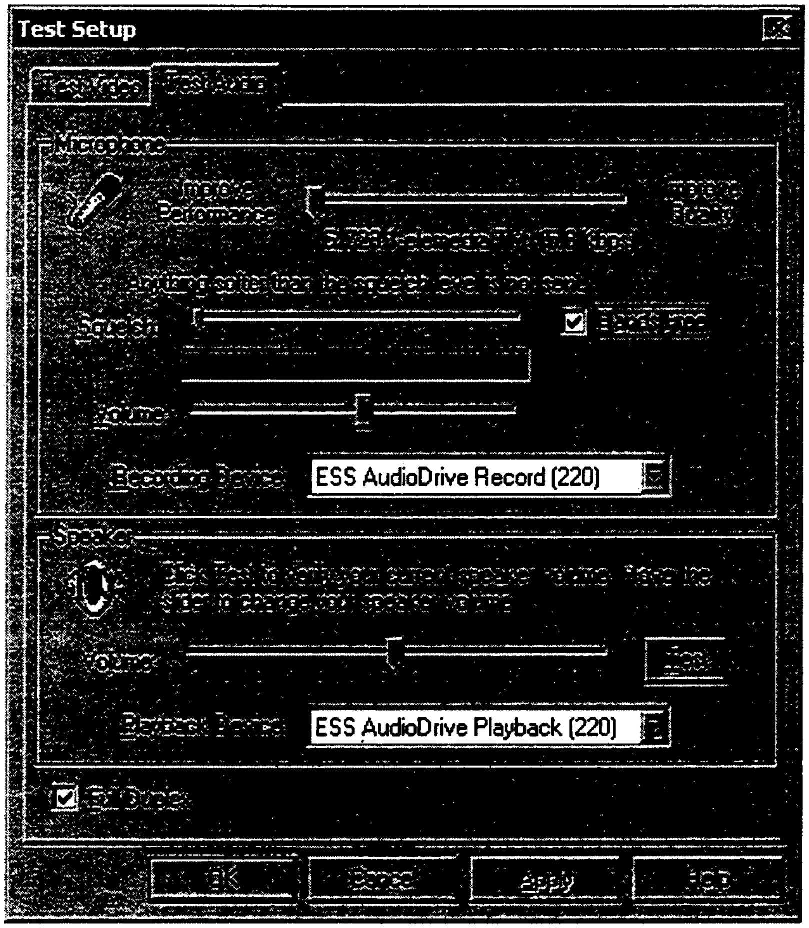


Figure 20: Audio Configuration for CU-SEE-ME

The above figure displays the audio configuration options. The G.723.1 codec is the version offered by Lucent Technologies. Both 5.3 and 6.4 Kbps version are offered. CU-SEE-ME is a commercial product offered by White Pine Software Inc.

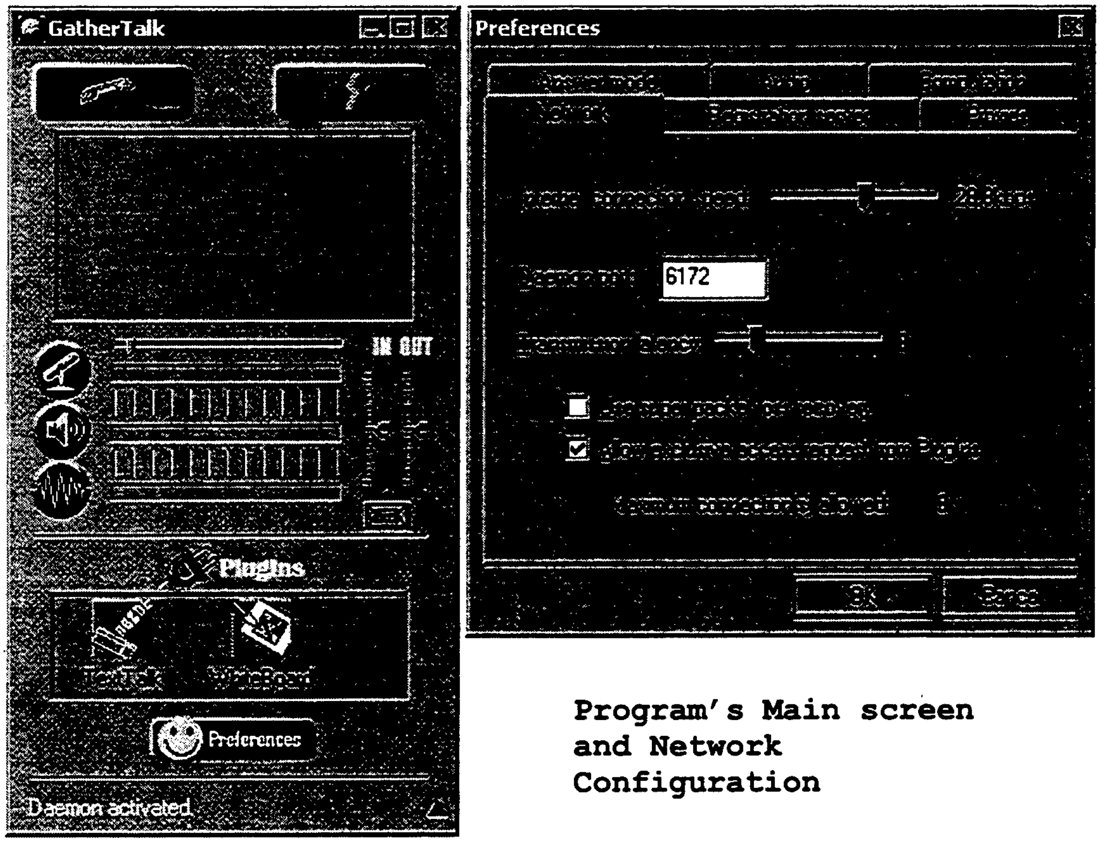
Table 10: CU-SEE-ME Summary of Results

Product name and Version	CU-SEE-ME 3.1.2 (Build 7)
Company or implementor	White Pine Software Inc.
LAN quality	8.5/10
Internet Quality	6.5/10
H.323 compliance	YES
Audio Compression algorithms	- G.723.1 (5.3 & 6.4 Kbps) - DigiTalk (8.5 Kbps) - Delta-Mod (16 Kbps) - Intel DVI (32 Kbps)
Interoperability	- Intel Video Phone - NetMeeting
Platforms	- Windows 95/98/NT 4.0 - MacOS
Address formats	- IP-Address - e-mail - DLS alias
Video support	YES
Other features	- Whiteboard

5. GatherTalk

GatherTalk research was initially performed under the project Interactive Voice Communications Systems. It was developed by Department of Electronic Engineering, the

Chinese University of Hong Kong. Starting from July of 1996, the Center for Internet Exchange Technologies of CUHK continued the development of GatherTalk. GatherTalk allows multi-party voice conferencing on the Internet, for 3-5 people.



Program's Main screen and Network Configuration

Figure 21: GatherTalk Screenshots

The voice quality was very good with no significant latencies or distortions. However, due to the proprietary algorithms used no interoperability with other products was achieved. The allowed user configuration is shown in figure

17. No manual choice for compression algorithms is allowed but obviously the appropriate codec is used when the user inputs transmission speed and latency.

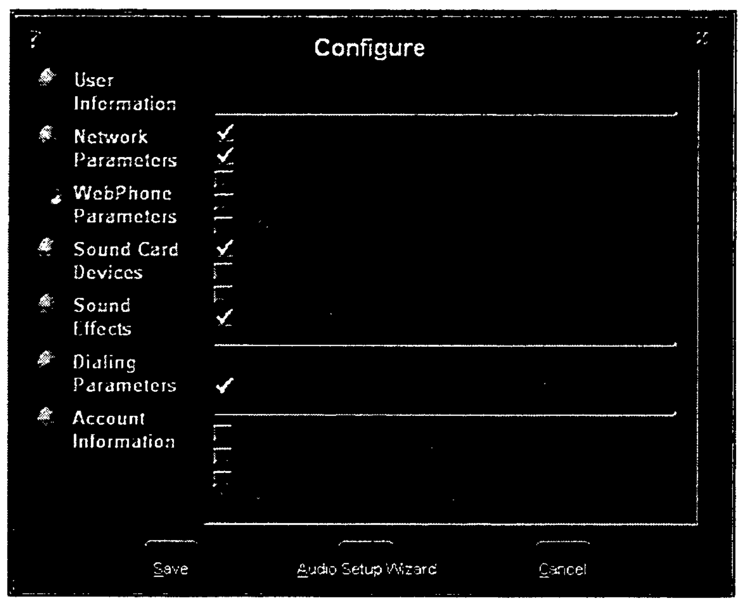
Table 11: GatherTalk Summary of Results

Product name and Version	GatherTalk 1.6
Company or implementor	The Chinese University of Hong Kong
LAN quality	8/10
Internet Quality	7/10
H.323 compliance	NO
Audio Compression algorithms	Proprietary
Interoperability	None
Platforms	- Windows 95/98/NT 4.0
Address formats	- IP-Address - e-mail
Video support	YES
Other features	- Whiteboard - Text Chat

6. WebPhone

WebPhone is one of the few to offer all the available Telephony and IP Telephony features over a LAN or the Internet including caller identification, multiple lines,

call forwarding, call transferring etc. The interface of the program resembles that of a standard telephone as shown in figure 18. The program is H.323 compliant and one of the few that has some interoperability. The voice quality was average with some latencies and distortion. The program allows the user to select H.323 compliance but does not include any manual selection for codecs.



Program's main panel and configuration (including H.323 option)

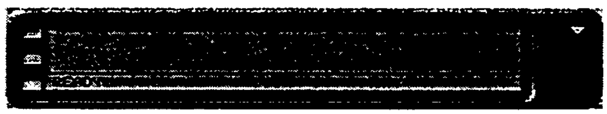
Figure 22: WebPhone Screenshots

Table 12: WebPhone Summary of Results

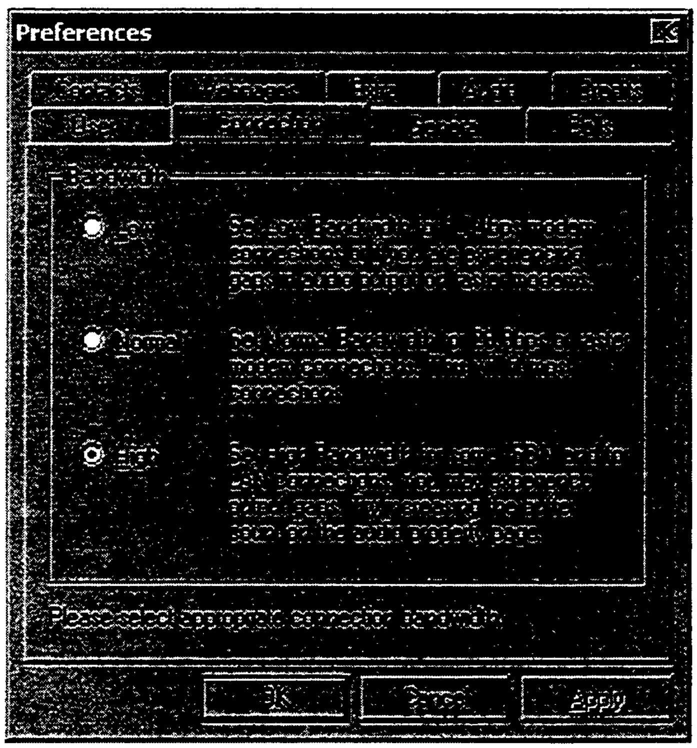
Product name and Version	WebPhone 4.02
Company or implementor	NetSpeak
LAN quality	7/10
Internet Quality	5.5/10
H.323 compliance	YES
Audio Compression algorithms	Proprietary
Interoperability	- NetMeeting - Intel Video Phone
Platforms	- Windows 95/98/NT 4.0
Address formats	- IP-Address - e-mail - Telephone number
Video support	YES
Other features	- Whiteboard - Text Chat - Call Transfer - Call blocking - Call conferencing - Caller Identification - Multiple Lines - On-hold music - Voice mail

7. BuddyPhone

This program has the simplest interface but it also allows minimal configuration. It is not H.323 compliant and thus could not communicate with any other program in the tests. The voice quality was good though both on LAN and the Internet.



Simple main panel



Configuration menu

Figure 23: BuddyPhone Screenshots

Table 13: BuddyPhone Summary of Results

Product name and Version	BuddyPhone 1.54
Company or implementor	Henry Pfluger
LAN quality	7/10
Internet Quality	6/10
H.323 compliance	NO
Audio Compression algorithms	Proprietary
Interoperability	None
Platforms	- Windows 95/98/NT 4.0
Address formats	- IP-Address - e-mail
Video support	NO
Other features	- Text Messages - Random calls

8. PGPfone

The PGPfone is short for the "Pretty Good Privacy Phone". It uses the encryption algorithms and keys used by the well-known PGP. The encryption capability is the uniqueness of the program and the one feature that differentiates it from the rest. The call setup involves negotiation for the proper choice of encryption algorithms. This adds a certain amount of overehead. Thus the program

behaved poorly in actual conversations both on LAN and on the Internet. The conversations would also terminate abruptly in the middle of a phone call.

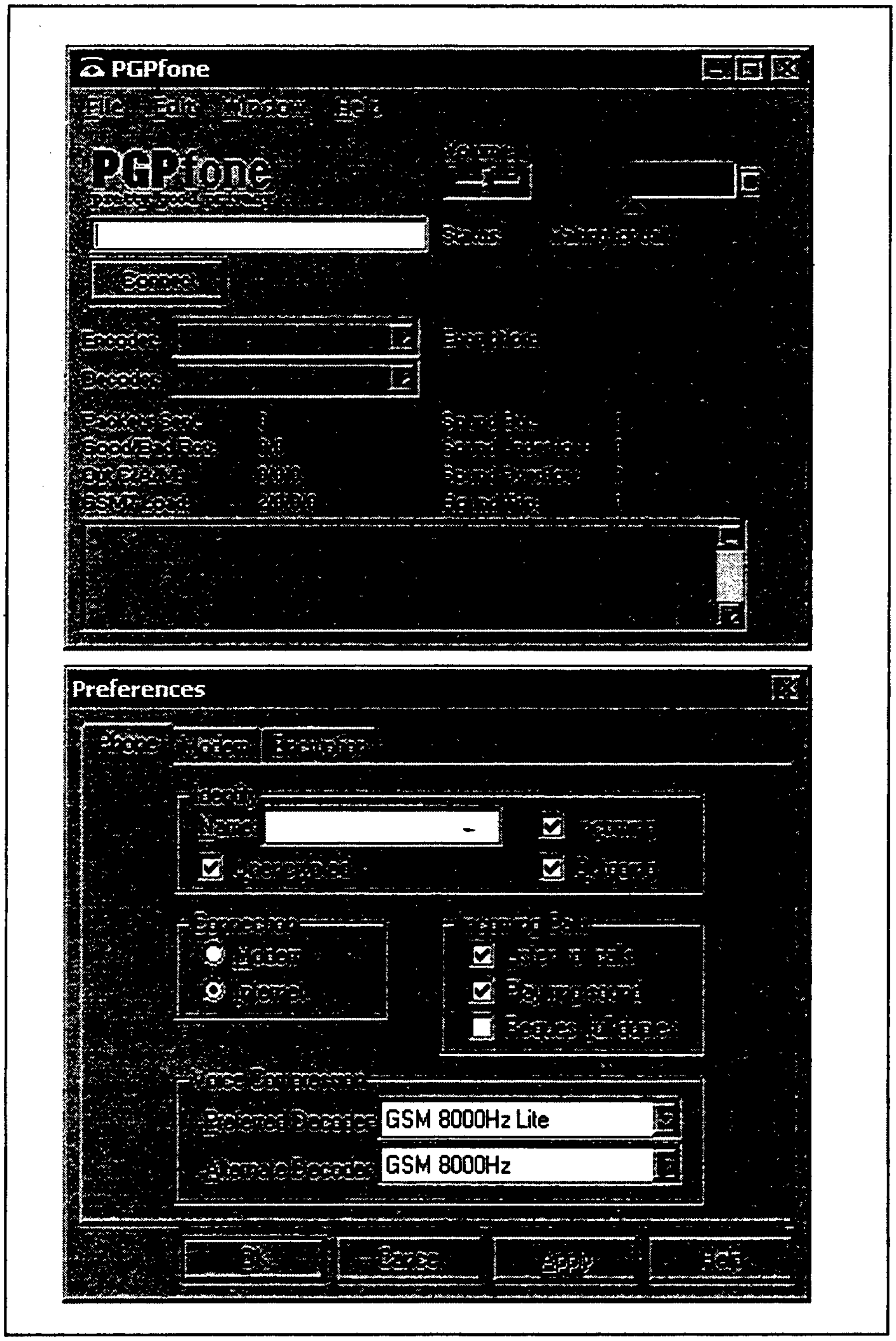


Figure 24: PGPfone Screenshots

The program uses proprietary protocols and supports a variety of GSM codec variations. Naturally it could not communicate with any other program. It allows extensive amount of configuration including manual choice of codecs and encryption algorithms (Blowfish, CAST, and TripleDES).

Table 14: PGPfone Summary of Results

Product name and Version	PGPfone 1.0b2
Company or implementor	Pretty Good Privacy, Inc.
LAN quality	5/10
Internet Quality	4/10
H.323 compliance	NO
Audio Compression algorithms	- GSM (4410 - 11025 Hz) - ADPCM (8 kHz)
Interoperability	None
Platforms	- Windows 95/98/NT 4.0 - MacOS
Address formats	- IP-Address
Video support	NO
Other features	- Encryption

9. Speak Freely

This simple program was the surprise of all the tests. Speak Freely is a pure audio Telephony application with the

simplest interface; no fancy toolbars or nicely designed buttons. However, the voice quality that this small program produced was unbeatable. The program uses three alternative protocols (RTP, and two proprietary ones), and a variety of codecs including GSM and LPC. The source code is freely available for any developer since it is an "Open Source Software" application. Currently the program comes in two versions; one for Windows and the other one for UNIX-based systems including Linux. The UNIX version is somewhat more primitive since it does not support graphical interface yet but both versions can communicate with each other. The developers of the program promise to enhance the UNIX version and they also intend to include H.323 support using an open source H.323 stack that is currently available.

The program allows the user to make extensive configurations including manual choice of codecs. Also the program includes an option which allows the user to select the desired level of jitter compensation, starting from none up to 3 seconds. The creators of the program have set up an echo server, which allows for individual testing without the need of a second calling party. The user connects to the server and transmits a voice message. The

message echoes back after approximately 10 seconds. Finally the program has encryption capabilities.

The quality of conversations and the fact that the source code is free for everyone brings this program in the first position despite the lack of Video capabilities and a more advanced interface. During the tests several combinations of choices were selected. The GSM codec produced the best results. The LPC produced somewhat more unnatural voice since it is a vocoder algorithm. However in both cases the voice was crispy-clear the losses were kept in a minor level and the latency was satisfactory. Overall an excellent program which deserves credit. The future enhancements will definitely make this program an unbeatable competitor.

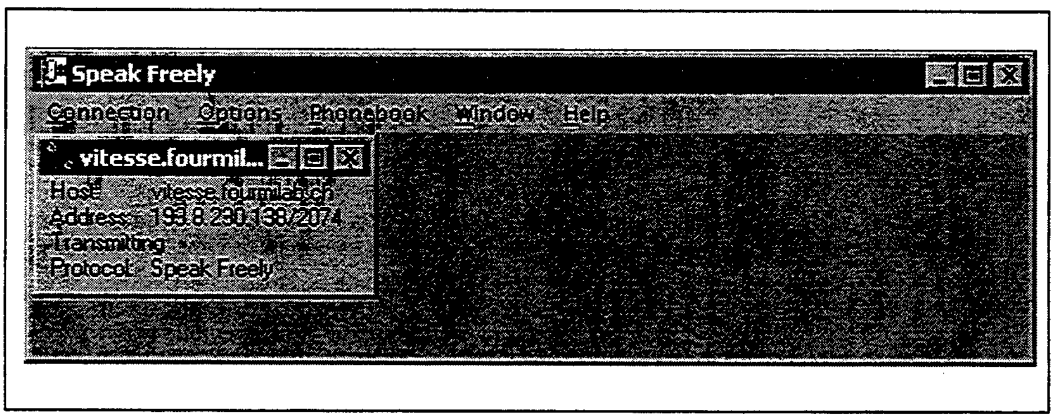


Figure 25: Speak Freely Screenshots

Table 15: Speak Freely Summary of Results

Product name and Version	Speak Freely 7.0
Company or implementor	John Walker, Brian Wiles
LAN quality	9/10
Internet Quality	8/10
H.323 compliance	NO
Audio Compression algorithms	- GSM - ADPCM - LPC - LPC-10
Interoperability	None
Platforms	- Windows 95/98/NT 4.0 - UNIX - Linux
Address formats	- IP-Address
Video support	NO
Other features	- Encryption - Answering Machine

10. Iris Phone

This was the last program of our tests, which produced satisfactory results. The program uses proprietary codecs for a variety of connections and allows the user to configure the program depending on the type of network connection. It is not H.323 compliant and it could not

communicate with any other program. The quality of voice was good on LAN but somewhat poorer on the Internet. The program uses a browser-like interface similar to the one used by CU-SEE-ME but it also has the option for a secondary "futuristic" interface. A decent program but it could be better.

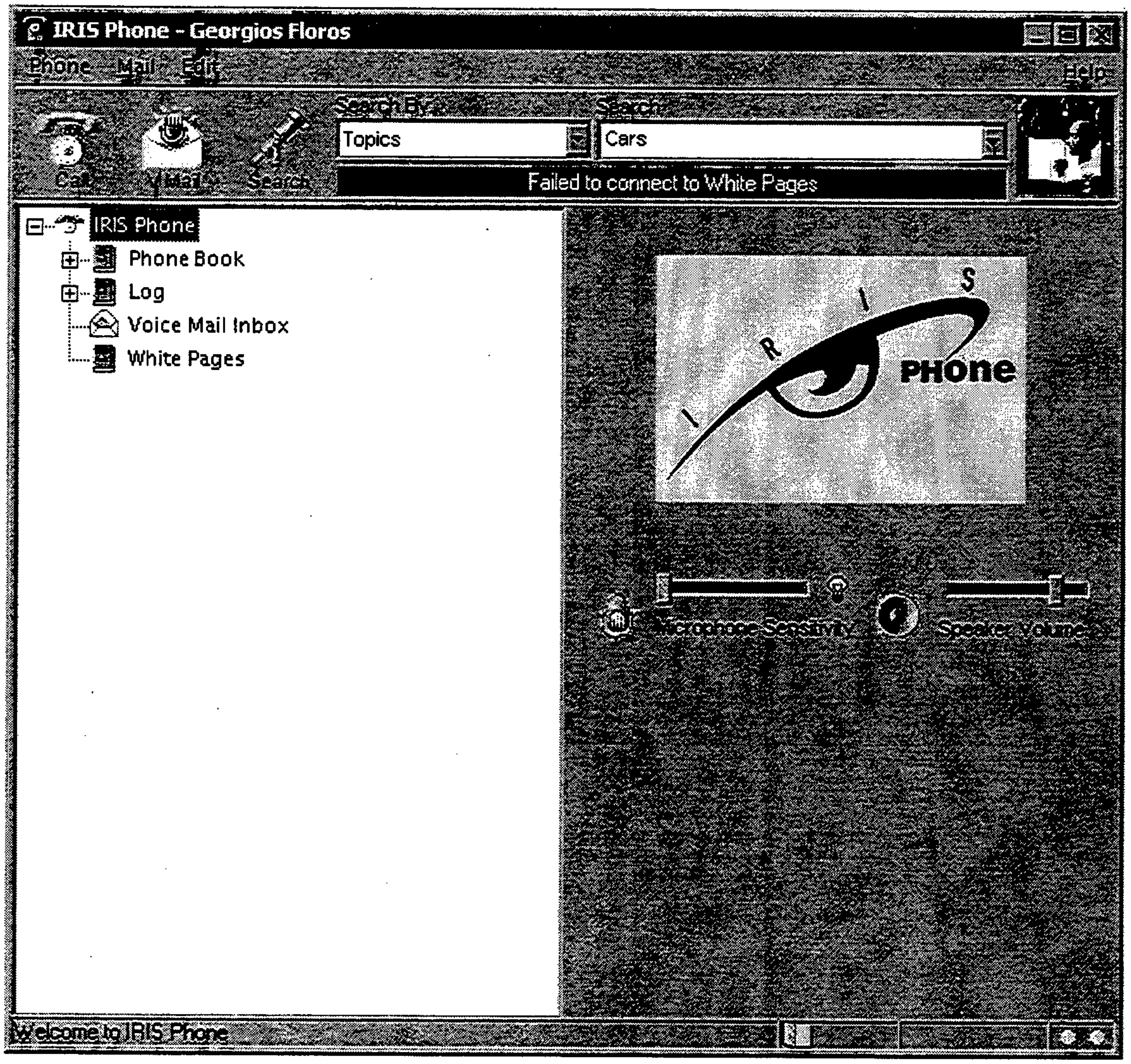


Figure 26: IRIS Phone Screenshot

Table 16: IRIS Phone Summary of Results

Product name and Version	IRIS Phone 3.0
Company or implementor	IRIS Systems
LAN quality	7/10
Internet Quality	5/10
H.323 compliance	NO
Audio Compression algorithms	- IRIS Audio Codecs (Proprietary)
Interoperability	None
Platforms	- Windows 95/98/NT 4.0
Address formats	- IP-Address - DLS alias
Video support	YES
Other features	- Text Chat

C. SUMMARY OF RESULTS

The majority of the programs performed relatively well on the LAN connection, which shows that these programs are ready for decent conversations on a small LAN. However the Internet results were rather poor with few exceptions. Half of the products were H.323 compliant. However as it was expected H.323 compliance does not guarantee interoperability. Internet Phone in particular, which

claims to be H.323 compliant, could not communicate with any other program including those that supported H.323. This can only mean two things. Either some vendors do not interpret the H.323 specifications correctly or the H.323 standard does not have strict requirements. The latter is probably the more likely reason.

As it was expected bandwidth was the main factor that affected program performance. Other factors like delay and jitter seem to be well addressed by the existing protocols and compression algorithms. Latency was a significant drawback in all the programs but this is a factor, which is implementation and protocol independent. The test also revealed that all programs prefer Windows 9X or NT as their platform while only a few supported other operating systems.

Most of the programs were highly configurable and provided options for different Network conditions. Video support was also a favorite feature but it severed audio quality due to the large bandwidth requirements. Thus the best audio quality came from a pure audio application which also happened to be open source software one. The results of the test are displayed in the following comparative charts and diagrams.

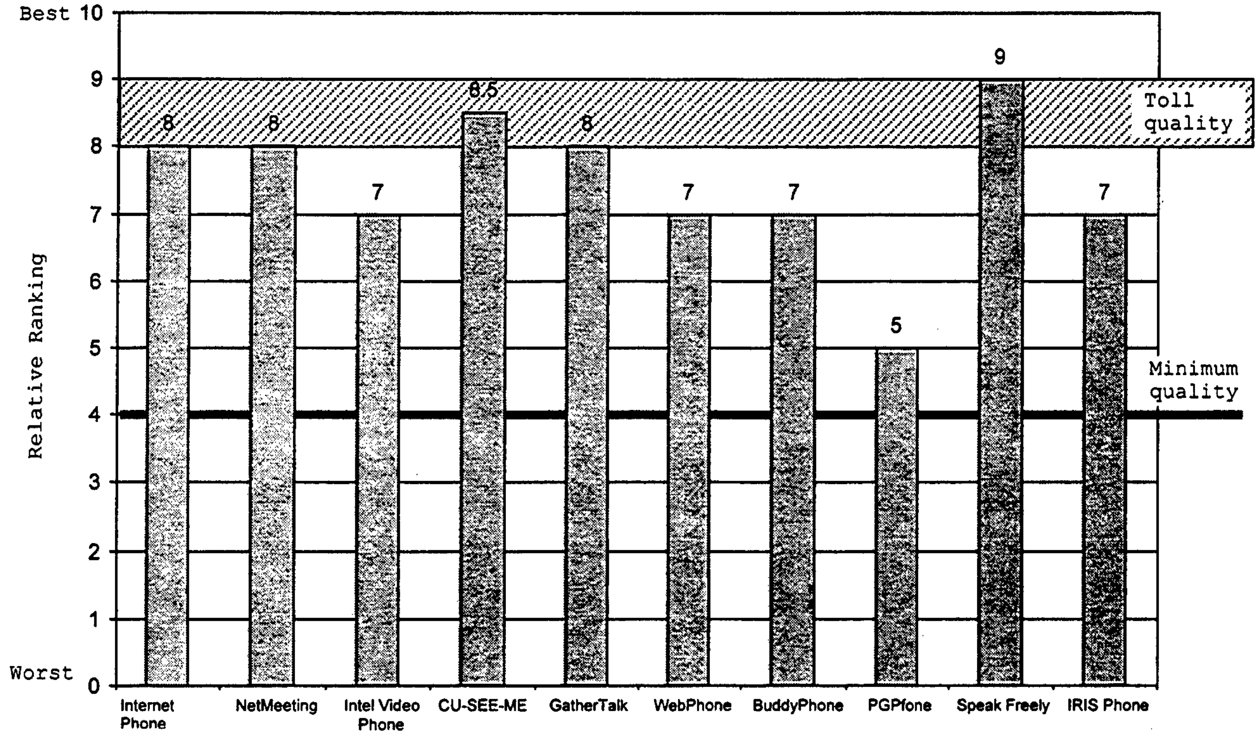


Figure 27: Voice Quality over LAN

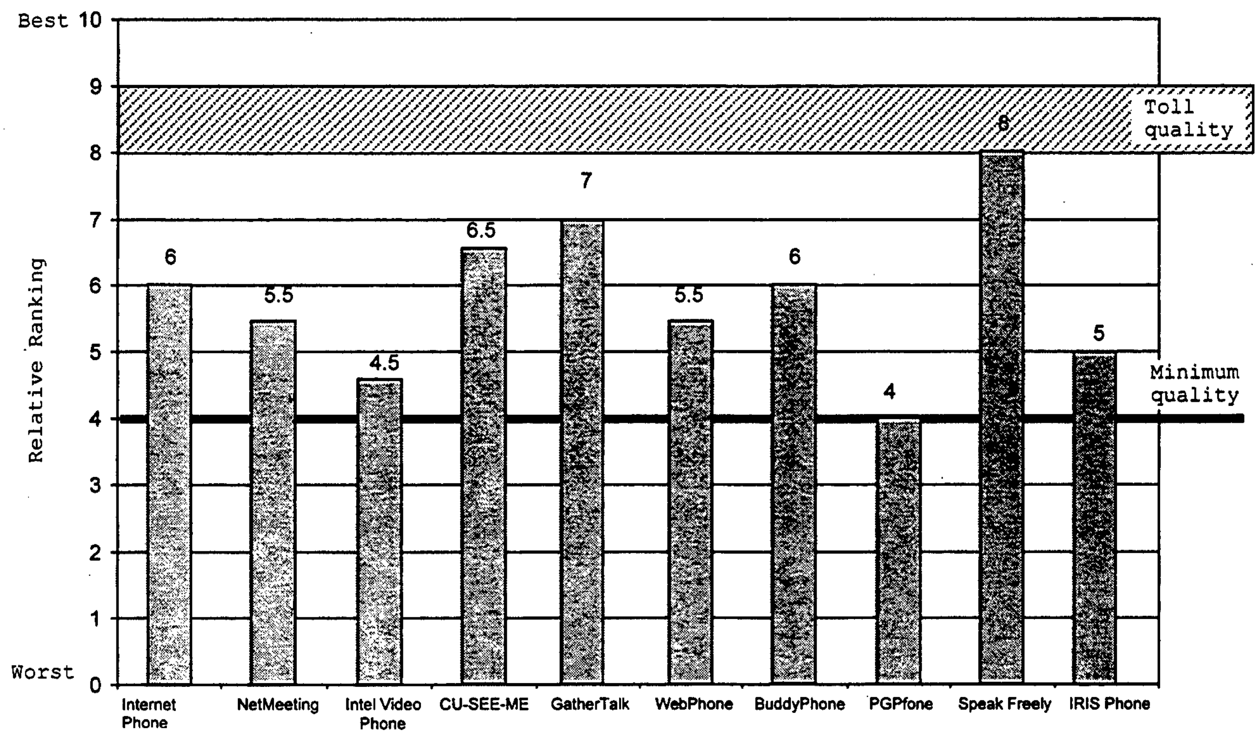


Figure 28: Voice Quality over the Internet

	Internet Phone	NetMeeting	CU-SEE-ME	Intel Video Phone	GatherTalk	WebPhone	BuddyPhone	PGPfone	Speak Freely	Iris Phone
Internet Phone										
NetMeeting			X	X		X				
CU-SEE-ME										
Intel Video Phone		X	X			X				
GatherTalk										
WebPhone		X		X						
BuddyPhone										
PGPfone										
Speak Freely										
Iris Phone										

Figure 29: Software Interoperability Chart

	Win 9X	Win NT	MacOS	UNIX	Linux
Internet Phone	X	X	X		
NetMeeting	X	X	X		
CU-SEE-ME	X	X			
Intel Video Phone	X	X			
GatherTalk	X	X			
WebPhone	X	X			
BuddyPhone	X	X			
PGPfone	X	X	X		
Speak Freely	X	X		X	X
Iris Phone	X	X			

Figure 30: Platform Compatibility Chart

VI. CONCLUSIONS AND RECOMMENDATIONS

A. CONCLUSIONS ON INTERNET TELEPHONY

Internet Telephony and Voice over IP in particular is viewed by some people to be an effective technology and by others as nothing more than an irritant. The irritating aspect stems from those people who have used the public Internet to make telephone calls. Indeed during our tests we always had the feeling that something was missing. There were very few occasions where our conversations could be directly compared with the quality of a standard long distance telephone call. However, it is the author's belief that we are about to see a tremendous expansion of this new technology in the next five to ten years.

One major reason for that is the universal presence of IP. The existence of IP in user personal computers and workstations gives IP a decided advantage over other existing technologies that are not resident in the user appliance. This "location" of IP makes it a very convenient platform from which to launch voice traffic. Moreover with IPv6 on the way Internet Telephony will find at last a competent and attractive ally.

Another significant reason is the maturation of technologies (or even better the maturation of expectations and demand). A few years ago Internet Telephony was just a "gimmick". Today this concept has attracted a lot of major computer, networking, and telecommunications companies. Names like *Microsoft, Intel, Sun, IBM, Siemens, Lucent, Cisco,* are involved one way or another with the Internet Telephony Market. Other companies are indirectly involved by supporting Internet telephony products. There is an exploding growth in data network investment. It is a fact that the world is experiencing a shift away from circuit-based networks to packet-based networks. Some market forecasts place the ratio of data networks-to-circuit networks at **80** to **20** percent by **2005**.

The fact that the tested applications performed so well over LAN indicates that this technology is ready to replace existing traditional PBX's. The quality of codecs is more than satisfactory even with low bandwidth. Long delays and large RTT's are still a problem but the data rates on the Internet are getting improved day by day. However, the deployment of Internet Telephony is not a trivial matter; but make no mistake, this technology is here to stay.

B. CONCLUSIONS AND RECOMMENDATIONS FOR H.323

In chapter II we mentioned that if Internet Telephony is to be succesfully deployed standardization is critical. In chapters III and IV we dissected the H.323 multimedia standard which is the proposed standard for Internet Telephony. It was clearly shown that the standard is not yet mature to achieve its goals. The standard is very complex and does not provide strict guidance and specification for the Internet Telephony implementors. However our tests showed that **5** out of **10** programs tested already supported the standard. Unfortunately H.323 compliance does not necessarily guarantee interoperability between products from different vendors. Only **4** out of **5** H.323 compliant products could communicate with each other. Last but not least H.323 compliance does not guarantee voice quality. The best voice quality was achieved with an "open source software" product that used proprietary protocols combined with RTP.

The nature of H.323 makes it more suitable for Local Area Networks (the initial target domain of the standard). However, it is apparent that the Internet Telephony industry is looking desperately for quick solution in the

standardization issue, and H.323 has gained significant popularity.

In order for the H.323 to be more suitable for Internet Telephony the author of this thesis recommends the following:

- The G.723.1 codec should become the default codec for Internet Telephony applications in place of G.711. The low bitrate (**5.3** and **6.4** Kbps) make the codec more suitable for low bandwidth connections.
- The standard must include more codecs in its specification. Currently only ITU codecs are included. There exist many attractive codecs (some of them in the public domain) which produced excellent results during our tests. The GSM codec (at **13** kBps) is one of them. The FS-1015 (LPC) is another.
- The call setup procedure has to be simplified. Therefore, The "Fast-Start" option must become mandatory and default for all H.323 implementations.
- A new section concerning the application of the standard over IP Networks has to be added in the appendices of the specification. H.323 may be

independent of the underlying layer; however, the unique features and constraints of Internet, which is based on IP, have to be addressed separately.

- The standard must provide detailed guidance on the allowed buffering size in the receiver endpoint. Currently recommendation H.245 allows the receiver to specify the maximum bitrate that may be sent but there is no further mentioning the maximum RTP packet size. Since delay and jitter can be resolved with buffering all Internet Telephony products use their own proprietary solutions for buffering thus creating interoperability problems.
- The standard has to provide more detailed instructions for security especially for the implementation of H.323 entities which will operate behind firewalls.

H.323 tends to become the single standards-based solution for a complete array of communication systems from simple point-to-point telephony to a rich multimedia conference with data sharing. Considering the amount of effort that is constantly being put for the development of standards we should expect the Internet Telephony industry

to converge on a solution which will combine H.323 with other alternatives like SIP.

LIST OF REFERENCES

Anonymous, (1998). A Primer on the H.323 Series Standard. Databeam Corporation White Paper.

Anonymous, (1998). The problems and pitfalls of getting H.323 safely through firewalls. Developer note, Intel Corporation.

Bernier, P. (1998). The Standards Struggle. Will SIP be a drain of H.323's momentum? Sounding Board Magazine.

Black, U. (1999). Voice Over IP. Prentice Hall.

Concalves, M. (1998) Voice over IP Networks. McGraw-Hill.

Held, G. (1998). Voice Over Data Networks. McGraw-Hill.

Hirschman, E. (1999). Why H.323 Compliance is not enough for carrier products. Internet Telephony Magazine.

ITU-T Recommendation H.323, (1998). Packet Based Multimedia Communications Systems. ITU.

Jain, R. (1999). Voice Over IP. Issues and challenges. The Ohio State University.

Margulies E. (1996). Understanding the Voice-Enabled Internet. Flatiron Publishing, Inc.

Minoli, D. and Minoli, E. (1998) Delivering Voice over IP Networks. John Wiley & Sons.

Raghavan, S. V. and Tripathi, S. K. (1998) Networked Multimedia Systems. Concepts, Architecture & Design. Prentice Hall.

Schulzrinne, H. and Rosenberg, J. (1998) A Comparison of SIP and H.323 for Internet Telephony.

Schulzrinne, H. and Rosenberg, J. (1998) Internet Telephony: Architecture and Protocols an IETF Perspective.

Stallings, W. (1998) High Speed Networks: TCP/IP and ATM Design Principles. Prentice Hall.

Toga, J. and ElGebaly, H. (1998) Demystifying Multimedia Conferencing Over the Internet Using the H.323 Set of Standards. Intel Technology Journal.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center 2
   8725 John J. Kingman Rd., Suite 0944
   Ft. Belvoir, VA 22060-6218

2. Dudley Knox Library 2
   Naval Postgraduate School
   411 Dyer Rd.
   Monterey, CA 93943-5101

3. Prof. Dan C. Boger, Code CS 1
   Chairman, Department of Computer Science
   Naval Postgraduate School
   Monterey, CA 93943-5106

4. Prof. Gilbert M. Lundy, Code CS/Ln 2
   Department of Computer Science
   Naval Postgraduate School
   Monterey, CA 93943-5106

5. Prof. Geoffrey C. Xie, Code CS/Xi 2
   Department of Computer Science
   Naval Postgraduate School
   Monterey, CA 93943-5106

6. Deborah Walsh, Code 01B1 1
   Distance Learning Coordinator
   Naval Postgraduate School
   Monterey, CA 93943-5106

7. Georgios Floros 2
   Sarantaporou 7 St.
   Athens 11144, GREECE